



Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

## REFERENCE

AUTHORS Kouchi,Y., Masago,A. and Takahata,T.  
TITLE Gene assay method for predicting glaucoma onset risk  
JOURNAL Patent: EP 1388590-A 5 11-FEB-2004;  
Sysmex Corporation (JP)

## FEATURES

source  
1..1074  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

## ORIGIN

Query Match 100.0%; Score 20; DB 6; Length 1074;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACGCCTT 20

Db 261 GAAACTGACCTTCACGCCTT 280

## RESULT 3

AF283519S5 AF283519S5 4015 bp DNA linear PRI 17-AUG-2000  
LOCUS Homo sapiens FIP2 gene, exons 7 and 8.  
DEFINITION AF283523  
ACCESSION AF283523  
VERSION AF283523.1 GI:9837251  
KEYWORDS  
SEGMENT  
5 of 9  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

REFERENCE 1 (bases 1 to 4015)

AUTHORS Li,D. and Roberts,R.  
TITLE Human FIP-2: Genomic structure and mutational analysis in ARVD patients

## JOURNAL

REFERENCE 2 (bases 1 to 4015)  
Unpublished  
AUTHORS Li,D. and Roberts,R.

TITLE Direct Submission  
JOURNAL Submitted (30-JUN-2000) Medicine, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

## FEATURES

Location/Qualifiers  
1..4015  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="10"

exon 853..926

exon 3477..3629

/number=7

/number=8

## ORIGIN

Query Match 100.0%; Score 20; DB 8; Length 4015;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACGCCTT 20

Db 613 GAAACTGACCTTCACGCCTT 632

## RESULT 4

AL355355 196203 bp DNA linear PRI 18-MAY-2005  
LOCUS Human DNA sequence from clone RP11-730A19 on chromosome 10 Contains the 5' end of a novel gene (DKFZP761F241) (FLJ20925 FLJ38473), a ribosomal protein L5 (RPL5) pseudogene, the OPTN gene for

optineurin, a small nuclear ribonucleoprotein polypeptide G (SNRPG) pseudogene, a pseudogene similar to part of COX10 homolog, cytochrome c oxidase assembly protein heme A: farnesyltransferase (yeast) (COX10), a novel pseudogene (FLJ10648 KIAA1525), a ribosomal protein L36A (RPL36A) pseudogene, the 5' end of the MCM10 gene for MCM10 minichromosome maintenance deficient 10 (S. cerevisiae), a pseudogene similar to part of chromodomain protein Y chromosome 1 (CDY1) and two CpG islands, complete sequence.

## ACCESSION

AL355355

VERSION AL355355.25 GI:16972859

KEYWORDS HTG; CDY1; chromodomain; COX10; CpG island; DKFZP761F241; FLJ10648; FLJ20925; FLJ38473; KIAA1525; MCM10; optineurin; OPTN; RPL36A; RPL5; SNRPG.

## SOURCE

Homo sapiens (human)

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

## REFERENCE

1 (bases 1 to 196203)

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Almeida,J.  
Direct Submission  
Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk  
Clone requests: clonerequest@sanger.ac.uk  
On Nov 16, 2001 this sequence version replaced gi:16214585.  
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr10

RP11-730A19 is from the library RPC1-11.3 constructed by the group of Pieter de Jong. For further details see

http://www.chori.org/bacpac/home.htm

VECTOR: pBAC3.6

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: vegas@sanger.ac.uk

-----

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

## FEATURES

## source

Location/Qualifiers  
1..196203  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/clone="RP11-730A19"  
/clone\_lib="RPC1-11.3"

## misc\_feature

1

/note="Clone left end: RP11-730A19"

join(complement(123034..123180),

complement(122612..122837),complement(122053..122130),

complement(97396..97628),complement(73202..73469),

complement(21866..22040),

complement(AL353586.14:43768..45820))

/gene="RP11-347122.1"

/locus\_tag="RP11-347122.1-001"

join(complement(123034..123180),

complement(122612..122837),complement(122053..122130),

complement(97396..97628),complement(73202..73469),

complement(21866..22040),

complement(AL353586.14:43768..45820))

## mRNA











```

repeat_region 11941..12005
/rpt_family="B4"
repeat_region 12071..12125
/rpt_family="Alu"
repeat_region 12877..12954
/rpt_family="Alu"
repeat_region 15557..15693
/rpt_family="Alu"
repeat_region 16143..16272
/rpt_family="B4"
repeat_region 16299..16477
/rpt_family="B2"
repeat_region 16478..16667
/rpt_family="B2"
repeat_region 16668..16683
/rpt_family="B2"
repeat_region 16717..16737
/rpt_family="MaLR"
repeat_region 16738..16883
/rpt_family="Alu"
repeat_region 16893..17031
/rpt_family="Alu"
repeat_region 17036..17223
/rpt_family="B2"
trna 17041..17113
/product="tRNA-Asn"
/Note="likely pseudogene (HMM Sc=35.92 / Sec struct
Sc=-12.35)"
repeat_region 17265..17364
/rpt_family="Alu"
repeat_region 17570..17759
/rpt_family="B2"
repeat_region 17760..17966
/rpt_family="B2"
repeat_region 17993..18083
/rpt_family="Alu"
repeat_region 18102..18312
/rpt_family="B2"
repeat_region 18419..18611
/rpt_family="B2"
repeat_region 18628..18765
/rpt_family="Alu"
repeat_region 19225..19290
/rpt_family="B4"
repeat_region 19270..19343
/rpt_family="B2"
repeat_region 19383..19767
/rpt_family="L1"
repeat_region 19769..19830
/rpt_family="B2"
repeat_region 19832..20140
/rpt_family="L1"
repeat_region 21039..21185
/rpt_family="Alu"
repeat_region 22572..22745
/rpt_family="MaLR"
repeat_region 22708..22881
/rpt_family="B4"
repeat_region 22758..22889
/rpt_family="Alu"
repeat_region 22911..23137
/rpt_family="B2"
repeat_region 23155..23328
/rpt_family="B2"
trna 23160..23231
/product="tRNA-Ser"
/Note="likely pseudogene (HMM Sc=38.08 / Sec struct
Sc=-10.52)"
repeat_region 23401..23538
/rpt_family="Alu"

```

Query Match 87.0%; Score 17.4; DB 9; Length 137334;  
 Best Local Similarity 94.7%; Pred. No. 5.3e+02;

```

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GAAACTGACCTTCACGCCT 19
|||||
Db 74460 GAAACTGACCTTCACCT 74478
|||||

RESULT 8
AC118211 192713 bp DNA linear ROD 30-JAN-2004
Mus musculus chromosome 8, clone RP24-22305, complete sequence.
AC118211
AC118211.10 GI:41393436
HTG.
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE
Birren,B., Nusbaum,C. and Lander,E.
1 (bases 1 to 192713)
Mus musculus chromosome 8, clone RP24-22305
Unpublished
2 (bases 1 to 192713)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Collangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., DeAtellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kelle,C., LaRocque,K., Lamazares,R.,
Landers,T., Lehoczkv,J., Levine,R., Lindblad-Toh,K., Liu,G.,
McClean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,W., Mienga,V., McKernan,K., Meldrim,J., Meneus,L.,
Mihoval,T., Norman,C.H., O'Connor,T., Naylor,J., Nguyen,C., Nicol,R.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (14-APR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 192713)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cooke,P., Corum,B., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kelle,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., Maclean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Meneus,L., Mihoval,T., Mienga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (09-DEC-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

```

REFERENCE  
AUTHORS

4 (bases 1 to 192713)  
 Birren,B., Nuebaum,C., Lander,E., Abouelleil,A., Allen,N.,  
 Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,  
 Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,  
 Collamore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,  
 Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,  
 Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,  
 Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,  
 Hall,J., Horton,L., Hulme,W., Iliiev,I., Johnson,R., Jones,C.,  
 Kanat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,  
 Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,  
 MacDonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,  
 Meldrim,J., Menes,J., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,  
 Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,  
 O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,  
 Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
 Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,  
 Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,  
 Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,  
 Vassiliev,H., Venkataranan,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,  
 Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission  
 Submitted (30-JAN-2004) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Jan 30, 2004 this sequence version replaced gi:39573819.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE  
JOURNAL

Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu

## COMMENT

----- Project Information  
 Center project name: L21334  
 Center clone name: 223\_O\_5

FEATURES  
source

Location/Qualifiers  
 1..192713  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="8"  
 /map="8"  
 /clone="RP24-223O5"  
 /clone\_lib="RPCI-24 Male Mouse BAC"  
 1..3665  
 /note="wgs end extension  
 clone end:Sp6"  
 /rpt\_family="tRNA-Ala-GCY"  
 complement(140..219)  
 1668..1706  
 /rpt\_family="B4"  
 1755..1827  
 /rpt\_family="B4A"  
 complement(2101..2138)  
 /rpt\_family="B3A"  
 complement(2139..2278)  
 /rpt\_family="B1 MM"  
 complement(2279..2463)  
 /rpt\_family="B3A"  
 2899..2956  
 /rpt\_family="AT rich"  
 complement(3195..3304)  
 /rpt\_family="PB1D7"  
 3345..3459  
 /rpt\_family="B1 MM"  
 complement(3537..3581)  
 /rpt\_family="B2 Mm2"  
 3582..3645  
 /rpt\_family="B2 Mm2"  
 complement(3646..3771)  
 /rpt\_family="B2 Mm2"

misc\_feature  
 3666..3669  
 /note="clone boundary  
 clone end:SP6  
 site:MboI"  
 complement(3879..4072)  
 /rpt\_family="B2\_Mm2"  
 4117..4139  
 /rpt\_family="TG)n"  
 complement(4199..4464)  
 /rpt\_family="B4A"  
 4630..4751  
 /rpt\_family="PB1D7"  
 4883..5025  
 /rpt\_family="MIR"  
 5074..5277  
 /rpt\_family="ID B1"  
 complement(5607..5740)  
 /rpt\_family="B1\_MM"  
 5763..5798  
 /rpt\_family="B1F"  
 6291..6322  
 /rpt\_family="TA)n"  
 complement(6475..6780)  
 /rpt\_family="MLT1A0"  
 7519..7546  
 /rpt\_family="AT rich"  
 complement(7658..7780)  
 /rpt\_family="ID B1"  
 complement(7789..7939)  
 /rpt\_family="B1F"  
 complement(9273..9840)  
 /rpt\_family="L1M2"  
 10679..10718  
 /rpt\_family="AT rich"  
 complement(10735..10856)  
 /rpt\_family="B1\_MM"  
 11108..11130  
 /rpt\_family="GC rich"  
 complement(11145..11243)  
 /rpt\_family="B1F"  
 complement(11276..11374)  
 /rpt\_family="U6"  
 11578..11631  
 /rpt\_family="TA)n"  
 complement(11689..12152)  
 /rpt\_family="RMER15"  
 12154..12214  
 /rpt\_family="TTTTG)n"  
 complement(12216..12349)  
 /rpt\_family="B1\_MM"  
 12368..12514  
 /rpt\_family="B1 MM"  
 complement(12598..12917)  
 /rpt\_family="ORR1A2"  
 12988..13121  
 /rpt\_family="B1 MM"  
 complement(13233..13457)  
 /rpt\_family="ID B1"  
 complement(13496..13891)  
 /rpt\_family="MTA"  
 14093..14115  
 /rpt\_family="TTTTG)n"  
 complement(14137..14249)

Query Match 87.0%; Score 17.4; DB 9; Length 192713;  
 Best Local Similarity 94.7%; Pred. No. 5.4e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GAAACTGACCTTCAGCCT 19  
 |||||  
 DB 161221 GAAACTGACCTTCAGCCT 161239

## RESULT 9

## AL807395/c

## LOCUS

## DEFINITION

AL807395 201238 bp DNA linear ROD 09-FEB-2005  
 Mouse DNA sequence from clone Rp23-81p12 on chromosome 11. Contains the 5' end of the Tcn2 gene for transcobalamin 2, two novel genes, a glyceraldehyde-3-phosphate dehydrogenase (Gapd) pseudogene, the Pesi1 gene for BCR1 domain containing pescadillo homolog 1 (zebrafish), the Gcst gene for galactosylceramide sulfotransferase, a ubiquitin-conjugating enzyme E2L 3 (Ube2l3) pseudogene, the Sec14l4 and Sec14l2 genes for SEC14-like 4 (S. cerevisiae) and 2, the ortholog of human and rat SEC14-like 3 (S. cerevisiae) SEC14L3, a ribosomal protein L29 (Rpl29) pseudogene, a transformer 2 alpha homolog (Drosophila) (Tra2a) pseudogene and a CpG island, complete sequence.

## ACCESSION

## AL807395

## VERSION

## HTG: CpG island; Gapd; Gcst; Pesi1; Rpl29; Sec14l2; Sec14l3;

## KEYWORDS

## Sec14l4; Tcn2; Tra2a; transcobalamin; Ube2l3.

## SOURCE

## Mus musculus (house mouse)

## ORGANISM

## Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

## Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

## 1 (bases 1 to 201238)

## Matthews, L.

## Direct Submission

## Submitted (04-FEB-2005) Wellcome Trust Sanger Institute, Hinxton,

## Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk

## Clone requests: clonesrequest@sanger.ac.uk

## On Sep 27, 2002 this sequence version replaced gi:22798341.

## The following abbreviations are used to associate primary accession

## numbers given in the feature table with their source databases:

## Em., EMBL; Swi., SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information

## on the WORMPEP database can be found at

## http://www.sanger.ac.uk/Projects/C\_elegans/wormpep

## Genome Center

## Center: Wellcome Trust Sanger Institute

## Center code: SC

## Web site: http://www.sanger.ac.uk

## Contact: vegas@sanger.ac.uk

## -----

## This sequence was finished as follows unless otherwise noted: all

## regions were either double-stranded or sequenced with an alternate

## chemistry or covered by high quality data (i.e., phred quality &gt;=

## 30); an attempt was made to resolve all sequencing problems, such

## as compressions and repeats; all regions were covered by at least

## one subclone; and the assembly was confirmed by restriction digest,

## except on the rare occasion of the clone being a YAC.

## Sequence from the Mouse Genome Sequencing Consortium whole genome

## shotgun may have been used to confirm this sequence. Sequence data

## from the whole genome shotgun alone has only been used where it has

## a phred quality of at least 30

## Rp23-81p12 is from the RPCI-23 Mouse BAC Library

## constructed by the group of Pieter de Jong.

## For further details see http://www.chori.org/bacpac/home.htm

## VECTOR: pBACe3.6.

## FEATURES

## Location/Qualifiers

## 1..201238

## /organism="Mus musculus"

## /mol\_type="genomic DNA"

## /db\_xref="taxon:10090"

## /chromosomes="11"

## /clone="Rp23-81p12"

## /clone\_lib="RPCI-23"

## Join(complement(3022..3168), complement(2402..2596),

## complement(AL807241.5:14888..15080),

## complement(AL807241.5:13498..13667),

## complement(AL807241.5:12439..12600),

## complement(AL807241.5:11200..11372),

## complement(AL807241.5:10894..11080),

## complement(AL807241.5:9509..9674),

## complement(AL807241.5:6788..6903),

## complement(AL807241.5:4690..5207))

## /gene="Tcn2"

## /locus\_tag="RP23-453E10.1-002"

## /product="transcobalamin 2"

## /note="match: ESTs: AA220699.1 AA958998.1 BB561898.1

/gene="Tcn2"  
 /locus\_tag="RP23-453E10.1-001"  
 Join(complement(3022..3168), complement(2402..2596),  
 complement(AL807241.5:14888..15080),  
 complement(AL807241.5:13498..13667),  
 complement(AL807241.5:12439..12600),  
 complement(AL807241.5:11200..11372),  
 complement(AL807241.5:10894..11080),  
 complement(AL807241.5:9509..9674),  
 complement(AL807241.5:6788..6903),  
 complement(AL807241.5:4690..5207))  
 /gene="Tcn2"  
 /locus\_tag="RP23-453E10.1-001"  
 /product="transcobalamin 2"  
 /note="match: ESTs: AA986297.1 BB575349.1 BB607921.1  
 BB686702.1 BB852365.1 BB862014.1 BF152292.1 BF533221.1  
 BF577867.1 BF780863.1 BF783448.1 BG871086.1 BG873078.1  
 BG936672.1 BG972658.1 B1101447.1 B1101979.1 B1103255.1  
 B1103432.1 B1143403.1 B1154686.1 B1155612.1 B1332905.1  
 B1410359.1 B1647392.1 B1661716.1 B1684450.1 BM247737.1  
 BQ926262.1 BQ941586.1 BU841330.1 BY089481.1 BY175399.1  
 BY204520.1 BY227358.1 BY229685.1 BY340944.1 BY342689.1  
 CB589700.1 CB589907.1 CB954507.1  
 match: CDNAs: AF054810.1 AF090686.1 BC003720.1"  
 Join(complement(3022..3087), complement(2402..2593),  
 complement(AL807241.5:14888..15080),  
 complement(AL807241.5:13498..13667),  
 complement(AL807241.5:12439..12600),  
 complement(AL807241.5:11200..11372),  
 complement(AL807241.5:10894..11080),  
 complement(AL807241.5:9509..9674),  
 complement(AL807241.5:6788..6903),  
 complement(AL807241.5:4690..5207))  
 /gene="Tcn2"  
 /locus\_tag="RP23-453E10.1-005"  
 Join(complement(3022..3087), complement(2402..2593),  
 complement(AL807241.5:14888..15080),  
 complement(AL807241.5:13498..13667),  
 complement(AL807241.5:12439..12600),  
 complement(AL807241.5:11200..11372),  
 complement(AL807241.5:10894..11080),  
 complement(AL807241.5:9509..9674),  
 complement(AL807241.5:6788..6903),  
 complement(AL807241.5:4690..5207))  
 /gene="Tcn2"  
 /locus\_tag="RP23-453E10.1-005"  
 /product="transcobalamin 2"  
 Join(complement(2985..3053), complement(2402..2596),  
 complement(AL807241.5:14888..15080),  
 complement(AL807241.5:13498..13667),  
 complement(AL807241.5:12439..12600),  
 complement(AL807241.5:11200..11372),  
 complement(AL807241.5:10894..11080),  
 complement(AL807241.5:9509..9674),  
 complement(AL807241.5:6788..6903),  
 complement(AL807241.5:4690..5207))  
 /gene="Tcn2"  
 /locus\_tag="RP23-453E10.1-002"  
 Join(complement(2985..3053), complement(2402..2596),  
 complement(AL807241.5:14888..15080),  
 complement(AL807241.5:13498..13667),  
 complement(AL807241.5:12439..12600),  
 complement(AL807241.5:11200..11372),  
 complement(AL807241.5:10894..11080),  
 complement(AL807241.5:9509..9674),  
 complement(AL807241.5:6788..6903),  
 complement(AL807241.5:4690..5207))  
 /gene="Tcn2"  
 /locus\_tag="RP23-453E10.1-002"  
 /product="transcobalamin 2"  
 Join(complement(2985..3053), complement(2402..2596),  
 complement(AL807241.5:14888..15080),  
 complement(AL807241.5:13498..13667),  
 complement(AL807241.5:12439..12600),  
 complement(AL807241.5:11200..11372),  
 complement(AL807241.5:10894..11080),  
 complement(AL807241.5:9509..9674),  
 complement(AL807241.5:6788..6903),  
 complement(AL807241.5:4690..5207))  
 /gene="Tcn2"  
 /locus\_tag="RP23-453E10.1-002"  
 /product="transcobalamin 2"  
 Join(complement(2985..3053), complement(2402..2596),  
 complement(AL807241.5:14888..15080),  
 complement(AL807241.5:13498..13667),  
 complement(AL807241.5:12439..12600),  
 complement(AL807241.5:11200..11372),  
 complement(AL807241.5:10894..11080),  
 complement(AL807241.5:9509..9674),  
 complement(AL807241.5:6788..6903),  
 complement(AL807241.5:4690..5207))  
 /gene="Tcn2"  
 /locus\_tag="RP23-453E10.1-002"  
 /product="transcobalamin 2"  
 Join(complement(2985..3053), complement(2402..2596),  
 complement(AL807241.5:14888..15080),  
 complement(AL807241.5:13498..13667),  
 complement(AL807241.5:12439..12600),  
 complement(AL807241.5:11200..11372),  
 complement(AL807241.5:10894..11080),  
 complement(AL807241.5:9509..9674),  
 complement(AL807241.5:6788..6903),  
 complement(AL807241.5:4690..5207))  
 /gene="Tcn2"  
 /locus\_tag="RP23-453E10.1-002"  
 /product="transcobalamin 2"  
 Join(complement(2985..3053), complement(2402..2596),  
 complement(AL807241.5:14888..15080),  
 complement(AL807241.5:13498..13667),  
 complement(AL807241.5:12439..12600),  
 complement(AL807241.5:11200..11372),  
 complement(AL807241.5:10894..11080),  
 complement(AL807241.5:9509..9674),  
 complement(AL807241.5:6788..6903),  
 complement(AL807241.5:4690..5207))  
 /gene="Tcn2"  
 /locus\_tag="RP23-453E10.1-002"  
 /product="transcobalamin 2"  
 Join(complement(2985..3053), complement(2402..2596),  
 complement(AL807241.5:14888..15080),  
 complement(AL807241.5:13498..13667),  
 complement(AL807241.5:12439..12600),  
 complement(AL807241.5:11200..11372),  
 complement(AL807241.5:10894..11080),  
 complement(AL807241.5:9509..9674),  
 complement(AL807241.5:6788..6903),  
 complement(AL807241.5:4690..5207))  
 /gene="Tcn2"  
 /locus\_tag="RP23-453E10.1-002"  
 /product="transcobalamin 2"  
 Join(complement(2985..3053), complement(2402..2596),  
 complement(AL807241.5:14888..15080),  
 complement(AL807241.5:13498..13667),  
 complement(AL807241.5:12439..12600),  
 complement(AL807241.5:11200..11372),  
 complement(AL807241.5:10894..11080),  
 complement(AL807241.5:9509..9674),  
 complement(AL807241.5:6788..6903),  
 complement(AL807241.5:4690..5207))  
 /gene="Tcn2"  
 /locus\_tag="RP23-453E10.1-002"  
 /product="transcobalamin 2"  
 Join(complement(2985..3053), complement(2402..2596),  
 complement(AL807241.5:14888..15080),  
 complement(AL807241.5:13498..13667),  
 complement(AL807241.5:12439..12600),  
 complement(AL807241.5:11200..11372),  
 complement(AL807241.5:10894..11080),  
 complement(AL807241.5:9509..9674),  
 complement(AL807241.5:6788..6903),  
 complement(AL807241.5:4690..5207))  
 /gene="Tcn2"  
 /locus\_tag="RP23-453E10.1-002"  
 /product="transcobalamin 2"  
 Join(complement(2985..3053), complement(2402..2596),  
 complement(AL807241.5:14888..15080),  
 complement(AL807241.5:13498..13667),  
 complement(AL807241.5:12439..12600),  
 complement(AL807241.5:11200..11372),  
 complement(AL807241.5:10894..11080),  
 complement(AL807241.5:9509..9674),  
 complement(AL807241.5:6788..6903),  
 complement(AL807241.5:4690..5207))  
 /gene="Tcn2"  
 /locus\_tag="RP23-453E10.1-002"  
 /product="transcobalamin 2"  
 Join(complement(2985..3053), complement(2402..2596),  
 complement(AL807241.5:14888..15080),  
 complement(AL807241.5:13498..13667),  
 complement(AL807241.5:12439..12600),  
 complement(AL807241.5:11200..11372),  
 complement(AL807241.5:10894..11080),  
 complement(AL807241.5:9509..9674),  
 complement(AL807241.5:6788..6903),  
 complement(AL807241.5:4690..5207))  
 /gene="Tcn2"  
 /locus\_tag="RP23-453E10.1-002"  
 /product="transcobalamin 2"  
 Join(complement(2985..3053), complement(2402..2596),  
 complement(AL807241.5:14888..15080),  
 complement(AL807241.5:13498..13667),  
 complement(AL807241.5:12439..12600),  
 complement(AL807241.5:11200..11372),  
 complement(AL807241.5:10894..11080),  
 complement(AL807241.5:9509..9674),  
 complement(AL807241.5:6788..6903),  
 complement(AL807241.5:4690..5207))  
 /gene="Tcn2"  
 /locus\_tag="RP23-453E10.1-002"  
 /product="transcobalamin 2"  
 Join(complement(2985..3053), complement(2402..2596),  
 complement(AL807241.5:14888..15080),  
 complement(AL807241.5:13498..13667),  
 complement(AL807241.5:12439..12600),  
 complement(AL807241.5:11200..11372),  
 complement(AL807241.5:10894..11080),  
 complement(AL807241.5:9509..9674),  
 complement(AL807241.5:6788..6903),  
 complement(AL807241.5:4690..5207))  
 /gene="Tcn2"  
 /locus\_tag="RP23-453E10.1-002"  
 /product="transcobalamin 2"  
 Join(complement(2985..3053), complement(2402..2596),  
 complement(AL807241.5:14888..15080),  
 complement(AL807241.5:13498..13667),  
 complement(AL807241.5:12439..12600),  
 complement(AL807241.5:11200..11372),  
 complement(AL807241.5:10894..11080),  
 complement(AL807241.5:9509..9674),  
 complement(AL807241.5:6788..6903),  
 complement(AL807241.5:4690..5207))  
 /gene="Tcn2"  
 /locus\_tag="RP23-453E10.1-002"  
 /product="transcobalamin 2"  
 Join(complement(2985..3053), complement(2402..2596),  
 complement(AL807241.5:14888..15080),  
 complement(AL807241.5:13498..13667),  
 complement(AL807241.5:12439..12600),  
 complement(AL807241.5:11200..11372),  
 complement(AL807241.5:10894..11080),  
 complement(AL807241.5:9509..9674),  
 complement(AL807241.5:6788..6903),  
 complement(AL807241.5:4690..5207))  
 /gene="Tcn2"  
 /locus\_tag="RP23-453E10.1-002"  
 /product="transcobalamin 2"  
 Join(complement(2985..3053), complement(2402..2596),  
 complement(AL807241.5:14888..15080),  
 complement(AL807241.5:13498..13667),  
 complement(AL807241.5:12439..12600),  
 complement(AL807241.5:11200..11372),  
 complement(AL807241.5:10894..11080),  
 complement(AL807241.5:9509..9674),  
 complement(AL807241.5:6788..6903),  
 complement(AL807241.5:4690..5207))  
 /gene="Tcn2"  
 /locus\_tag="RP23-453E10.1-002"  
 /product="transcobalamin 2"  
 Join(complement(2985..3053), complement(2402..2596),  
 complement(AL807241.5:14888..15080),  
 complement(AL807241.5:13498..13667),  
 complement(AL807241.5:12439..12600),  
 complement(AL807241.5:11200..11372),  
 complement(AL807241.5:10894..11080),  
 complement(AL807241.5:9509..9674),  
 complement(AL807241.5:6788..6903),  
 complement(AL807241.5:4690..5207))  
 /gene="Tcn2"  
 /locus\_tag="RP23-453E10.1-002"  
 /product="transcobalamin 2"  
 Join(complement(2985..3053), complement(2402..2596),  
 complement(AL807241.5:14888..15080),  
 complement(AL807241.5:13498..13667),  
 complement(AL807241.5:12439..12600),  
 complement(AL807241.5:11200..11372),  
 complement(AL807241.5:10894..11080),  
 complement(AL807241.5:9509..9674),  
 complement(AL807241.5:6788..6903),  
 complement(AL807241.5:4690..5207))  
 /gene="Tcn2"  
 /locus\_tag="RP23-453E10.1-002"  
 /product="transcobalamin 2"  
 Join(complement(2985..3053), complement(2402..2596),  
 complement(AL807241.5:14888..15080),  
 complement(AL807241.5:13498..13667),  
 complement(AL807241.5:12439..12600),  
 complement(AL807241.5:11200..11372),  
 complement(AL807241.5:10894..11080),  
 complement(AL807241.5:9509..9674),  
 complement(AL807241.5:6788..6903),  
 complement(AL807241.5:4690..5207))  
 /gene="Tcn2"  
 /locus\_tag="RP23-453E10.1-002"  
 /product="transcobalamin 2"  
 Join(complement(2985..3053), complement(2402..2596),  
 complement(AL807241.5:14888..15080),  
 complement(AL807241.5:13498..13667),  
 complement(AL807241.5:12439..12600),  
 complement(AL807241.5:11200..11372),  
 complement(AL807241.5:10894..11080),  
 complement(AL807241.5:9509..9674),  
 complement(AL807241.5:6788..6903),  
 complement(AL807241.5:4690..5207))  
 /gene="Tcn2"  
 /locus\_tag="RP23-453E10.1-002"  
 /product="transcobalamin 2"  
 Join(complement(2985..3053), complement(2402..2596),  
 complement(AL807241.5:14888..15080),  
 complement(AL807241.5:13498..13667),  
 complement(AL807241.5:12439..12600),  
 complement(AL807241.5:11200..11372),  
 complement(AL807241.5:10894..11080),  
 complement(AL807241.5:9509..9674),  
 complement(AL807241.5:6788..6903),  
 complement(AL807241.5:4690..5207))  
 /gene="Tcn2"  
 /locus\_tag="RP23-453E10.1-002"  
 /product="transcobalamin 2"  
 Join(complement(2985..3053), complement(2402..2596),  
 complement(AL807241.5:14888..15080),  
 complement(AL807241.5:13498..13667),  
 complement(AL807241.5:12439..12600),  
 complement(AL807241.5:11200..11372),  
 complement(AL807241.5:10894..11080),  
 complement(AL807241.5:9509..9674),  
 complement(AL807241.5:6788..6903),  
 complement(AL807241.5:4690..5207))  
 /gene="Tcn2"  
 /locus\_tag="RP23-453E10.1-002"  
 /product="transcobalamin 2"  
 Join(complement(2985..3053), complement(2402..2596),  
 complement(AL807241.5:14888..15080),  
 complement(AL807241.5:13498..13667),  
 complement(AL807241.5:12439..12600),  
 complement(AL807241.5:11200..11372),  
 complement(AL807241.5:10894..11080),  
 complement(AL807241.5:9509..9674),  
 complement(AL807241.5:6788..6903),  
 complement(AL807241.5:4690..5207))  
 /gene="Tcn2"  
 /locus\_tag="RP23-453E10.1-002"  
 /product="transcobalamin 2"  
 Join(complement(2985..3053), complement(2402..2596),  
 complement(AL807241.5:14888..15080),  
 complement(AL807241.5:13498..13667),  
 complement(AL807241.5:12439..12600),  
 complement(AL807241.5:11200..11372),  
 complement(AL807241.5:10894..11080),  
 complement(AL807241.5:9509..9674),  
 complement(AL807241.5:6788..6903),  
 complement(AL807241.5:4690..5207))  
 /gene="Tcn2"  
 /locus\_tag="RP23-453E10.1-002"  
 /product="transcobalamin 2"  
 Join(complement(2985..3053), complement(2402..2596),  
 complement(AL807241.5:14888..15080),  
 complement(AL807241.5:13498..13667),  
 complement(AL807241.5:12439..12600),  
 complement(AL807241.5:11200..11372),  
 complement(AL807241.5:10894..11080),  
 complement(AL807241.5:9509..9674),  
 complement(AL807241.







```
/transl_table=11
/product="hypothetical protein"
/protein_id="AAO79252.1"
/db_xref="GI:29341465"
/translation="MLKOLLTVLLAIICLINVOAQLTPPAGTFFRLGISKGTDSHLA
POEKYKGAFAFKWLPDRGRFLEVAVTSLOADTLFWSFGNCQPDMDINVSVEGOA
FKCYGESMKLRLTAQVPTDDIRLSNGRQKTPILLYESGKRTDRPVLACRCLPAA
SKLYFCFEQNARADYNFMPLDFLFAKIDESHKSK"
complement(5111..5326)
/locus_tag="BT4148"
complement(5111..5326)
/locus_tag="BT4148"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAO79253.1"
/db_xref="GI:29341466"
/translation="WRLNFPYLYGIEAQTNKPGKICYTRNFNLSEITLIDVKDKDAIVL
KENQSNHFDFVKTITDRTTAGNLPH"
complement(5572..6630)
/locus_tag="BT4149"
complement(5572..6630)
/locus_tag="BT4149"
/codon_start=1
/transl_table=11
/product="exo-poly-alpha-D-galacturonosidase precursor"
/protein_id="AAO79254.1"
/db_xref="GI:29341467"
/translation="WRIKNLLPALLAGSLMIACITPAKCAAGNSFEMGQVPOQDLSWAD
SVGSLQPDGRMIIISANGFAGVADSTVISTEAIQKIDS CAVSGGGTVLQPGYQTG
ALFVSGNVLQKGVKTLASPDHHPFERSIAGIEMTPAAVINIVDKNAAISG
EGLDCRGKFWDKWNRKEYEARLWIVDYDCRVGLIVERSSDITLSGFTLMR
TGFWGQIDSYCTIDGLTINNIHGHPSTDGIDIDSSCNLILENDCVDNDNIC
IKSGRDADGLRVNRPTEENVNCTARKAGAGLITCGSETSGSIRNVGLDKAVGTYT
VLRKLSAMVGDPLRIS"
complement(6655..7893)
/locus_tag="BT4150"
complement(6655..7893)
/locus_tag="BT4150"
/codon_start=1
/transl_table=11
/product="putative rhamnogalacturonan acetyltransferase"
/protein_id="AAO79255.1"
/db_xref="GI:29341468"
/translation="WKTPLGLLLTATAISAEQARTFQADAPRYSEETGYGDLA
PTPEKSGRAPFVSVPDGNKVTVRLGSKKQAGVTVRGESRELVDNLPTRKGQF
TEETFIINKNRPIKSGSVRIKPREKTKLNWDDKLTLEFGDADFLCQISIEPADPS
VITFLCGNSTVVDQDNBFWASWGMQIPHFFGTVDVCIANIYESGESANTFICAKRLAK
ALSQIKGQDLYLPMFEGHNDQKQFGKAGAYSFMTSLKTFIDEARAGAHPLVTPPTQ
RRSFDANGHIRTDEYDPEAMRWLAAKENVPLIDLNMETRTLIEALGPTSKRAFPVY
PAGTYPGOTRNFADNTHNPYCAQIAQCQVIBGMKAVPELAKHLKIDPAYNPAHPDD
VNTFHNWSPTFIEKPDGN"
complement(7911..10787)
/locus_tag="BT4151"
complement(7911..10787)
/locus_tag="BT4151"
/codon_start=1
/transl_table=11
/product="beta-glucuronidase"
/protein_id="AAO79256.1"
/db_xref="GI:29341469"
/translation="WKOLLVSTAIIVPASVLITATGNHSHFSLSGKWDFOIDREDAGIKE
QWPNKSLDESINLPGSMEKLAGDNVARTQWTSGLYDSSYFENPYMEKYRLEGQVKL
PFFLPDRHYGVANTQKRTIPSDWKGRIILLFLERHETIETVWNTKEIGMQNSLC
VPHVYDLSAVTSKPCRTITRVDNRKEINVPDSSHITDQTQWNGIVGKIELQT
TPKVVYFDQIVPDLISHGKALVRMNVKSSSAKGEITLSAESFNTDTRHNVLPVQOTF
RIQAGNVRMELPMGKPLTFDEFNPALYKLTATLNSGKQSEKKQVQGRWDFKIEG
KWFYNGKRTMLRGTVENCDFPLTGVAPMDVASWERVRIENYGLNHRHSHFCFPE
AAPIAADLVGYLQPEGSWFNHGRNLGNQPIDKLMDETIALTKVEYNVASYCMLA
CNEPESRWVPWSKFDYWKAITDRHVVTYASVGSQWQMPQHNVHYKAGARGLSWA
GSPESMSDYSIVKQPYVSHETQWCAFPNFSERIKYTGTVGNKAVKGFIRDLIN
DNHMGSMGHDFFMWSGKLQAIICYKHEIKTLRTPYAGFQLIALNDYSGQGTALVGLL
DVFEFKGYINADEFRRCSPVPLARIKPFVYTNDDETTHADIEVSHFGCAAPLOCAKT
```

gene

CDS

gene

CDS

gene

CDS

gene

CDS

```
VYSIKDEYKGYAHGTVGTQNI PVGNLCPLGSDVMKLSGITRPOKLNNEIRIEGSDAV
NDWDFWVYPAQVELAQGNVYTTDLDAKAIISILEGGNVLTITAGAKIOYIGKEVQYFT
PVFWNTWFKMRPPHTTIFLNEYHPLFREFTTEYHNLQWMLLNKAQVQHQVQTFPPA
EFQOTISIDTWFINRKIGMFEANVLNGKLMTSMIDTTSKPEKRVARQWQHKAILEDY
MNSDAFRPTANIAPELIQELFTKVAGDVKS YTKDSPDELKPKIN"
11051..14011
/locus_tag="BT4152"
11051..14011
/locus_tag="BT4152"
/codon_start=1
/transl_table=11
/product="beta-galactosidase I"
/protein_id="AAO79257.1"
/db_xref="GI:29341470"
/translation="MALTLMSVMTTIFAGETAYLFYSYFINDSKOGHLAYSVDGLNWTPT
LNGRSFLAPSVGDKMLRDPISIQAPDGTFFHVMVTSWTDRIIIGYASSRDLIHWSQ
QAIIVMHPEBAHNCWAPELFYDEPSETYIIFWATTIPGRHKEYVPTSESEGLNHRMY
YVTTKDFRTSKTTFMFPNDFSVIDAALVKDPTQGDLMIVKNENSNPENKRLVTRT
DVSQVSFPKGRHGTAPVADSVVESLVDDRKHQSVKAQTSWPNFNDKELTLTVGYYY
PEHWDESQWERDPFKRMHELGEFTHFABFAWQLEPEGRYDFAWLDKAVLAAKYDL
KVIWCTSTATPVMMSRKYPEILLKNEGTVLDHGAROHSFASPLYRELSYKMTLKL
AOHYGNSRIVGQOLDNEPAVOFDYNPKAEQAFRDYLRKYNHNIQLLNDANGTAFWS
EAYSFPEITLPRKQVNMHHQILDYRPAQAQTNDFLNEQCLLIRKAKNQWITTN
YIPYDEHGIGSPALDFQSYTRYMVYVDNEGIRGRYRVGNPURIANWDFFRPIQG
TYGMELQPGQVNGWSINPQPLGAVRLMMWSVFAGSDFTCTYTRQPLXYGTEQHY
GIVGTGVTVTPGGREYETFIKEITRELKHYAPRETQKPADYLAARLTAILFNHNSWSI
ERQONRTWDTFAHIEKYRTLKSFQAPVDFVSEOKELTEYVVTIAPVQLADKSLVN
KWIAVYKNGNVLATCCTAOKDYGRGLPEAPFGSMITPLTGNEMNFYDLILPEDPCTV
VNMQKQEWNTWGEILIPASDSQWATYANEFYEGGPATFRKLKGKGVTVYGVVDSHN
GALKEDILKLYLAQIIPVMDLPYGITVEYRNLGLIVLNYSDRSYTFDIPELGSKVLVG
TEETPTAGLVYFSM"
14055..15353
/locus_tag="BT4153"
14055..15353
/locus_tag="BT4153"
/codon_start=1
/transl_table=11
/product="Exopolysaccharuronase precursor"
/protein_id="AAO79258.1"
/db_xref="GI:29341471"
/translation="WKQIVGLVLILCCTQWVAQVFPDGKAIPEWFRENKEVNVNALG
KKYSITDYGAVANDSTVVTQTKIQEVIDLAARNGGVIVIPKGTFLUGSLFFKNNTLHL
LEEDAVLKGSDDI SHFPVMTMRMEGOTLKYPMALVNDAGLDGFTISGRGLNGRLY
LNEADVLRKRSINPDCNTMEEMPRIIYILSNCKNTQI EGIRIMNSPFWSTHFYKCSFLK
LNLRTSPASPVKAPSTDAVDLVDCNNVLI KNCYMSVNDDAVALKGGKGPWADKANN
GENDIILEDCTYGFCHSAFTCGSESILSRNVIYRRCVKDRARALLTLKMSRDTQKY
EYILVEDTGNVQNCNLNAPWTQFDYDKRDKDTPLSYNNITMRNIDIDCNVFFNVKK
SDQYKLSDFCFENITVRAKKGKQVSDSFTMNNVKINQ"
```

Query Match

Best Local Similarity 87.0%; Score 17.4; DB 1; Length 300297;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAAACTGACCTTCACGCCT 19

Db 77533 GAAACTGACCTTCATGCCT 77515

RESULT 11

AKI73836/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

AKI73836 1456 bp mRNA linear INV 01-SEP-2004

Ciona intestinalis cDNA, clone:cic1017i03, full insert sequence.

AKI73836

AKI73836.1 GI:51774431

FLI CDNA.

Ciona intestinalis

Ciona intestinalis

Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;

Phlebobranchia; Cionidae; Ciona.

1

Satou, Y., Yamada, L., Mochizuki, Y., Takatori, N., Kawashima, T.,

Sasaki, A., Hamaguchi, M., Awazu, S., Yagi, K., Sasakura, Y.,

Nakayama, A., Ishikawa, H., Inaba, K. and Satoh, N.  
A cDNA resource from the basal chordate *Ciona intestinalis*  
Genesis 33 (4), 153-154 (2002)  
12203911  
2  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
PUBLISHED Only in Database (2004)  
3 (bases 1 to 1456)  
Satou, Y.  
Direct Submission  
Submitted (27-AUG-2004) Yutaka Satou, Kyoto University, Department  
of Zoology; Sakyo-ku, Kyoto, Kyoto 606-8502, Japan  
(E-mail: yutaka@ascidian.zool.kyoto-u.ac.jp, Tel: 81-75-753-4095,  
Fax: 81-75-705-1113)  
Ciona intestinalis cDNA Project (URL:  
http://ghost.zool.kyoto-u.ac.jp/indexr1.html).  
Location/Qualifiers  
1. 1456  
/organism="Ciona intestinalis"  
/mol\_type="mRNA"  
/db\_xref="taxon:7719"  
/clone="cic1017i03"  
ORIGIN  
Query Match 84.0%; Score 16.8; DB 2; Length 1456;  
Best Local Similarity 90.0%; Pred. No. 8.4e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 GAAACTGACCTTCACGCCTT 20  
||||| ||||||| |||||  
Db 933 GAAACGGACCTTCACGCTT 914  
RESULT 12  
AY062596 1570 bp mRNA linear PLN 25-NOV-2001  
LOCUS  
Arabidopsis thaliana heat shock protein 101 (At1g74310; F1017.2)  
mRNA, partial cds.  
AY062596  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Arabidopsis thaliana (thale cress)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 1570)  
Southwick, A., Karlin-Neumann, G., Nguyen, M., Lam, B., Miranda, M.,  
Palm, C.J., Bowser, L., Jones, T., Banh, J., Carninci, P., Chen, H.,  
Cheuk, R., Chung, M.K., Hayashizaki, Y., Ishida, J., Kamiya, A.,  
Kawai, J., Kim, C., Lin, J., Liu, S.X., Narusaka, M., Pham, P.K.,  
Sakano, H., Sakurai, T., Satou, M., Seki, M., Shinn, P., Yamada, K.,  
Shinozaki, K., Ecker, J., Theologis, A. and Davis, R.W.  
Direct Submission  
Submitted (14-NOV-2001) DNA Sequencing and Technology Center,  
Stanford University, 855 California Avenue, Palo Alto, CA 94304,  
USA  
e-mail for correspondence: arab@sequence.stanford.edu  
RIKEN Genomic Sciences Center (GSC) members carried out the  
collection and clustering of RAPL cDNAs (RAPL cDNA : 'RIKEN  
Arabidopsis Full-Length cDNA') : Seki, M., Narusaka, M., Ishida, J.,  
Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,  
Hayashizaki, Y. and Shinozaki, K.  
The Salk, Stanford, PGECC (SSP) Consortium members carried out the  
sequencing and annotation of the RAPL cDNAs: Nguyen, M., Palm, C.J.,  
Southwick, A., Karlin-Neumann, G., Lam, B., Miranda, M., Palm, C.J.,  
Bowser, L., Jones, T., Banh, J., Chen, H., Cheuk, R., Chung, M.K.,  
Kim, C., Lin, J., Liu, S.X., Pham, P.K., Sakano, H., Shinn, P.,  
Yamada, K., Ecker, J., Theologis, A. and Davis, R.W.

Southwick, A. (SSP/Stanford) and Seki, M. (RIKEN GSC) contributed  
equally to this work. Shinozaki, K. (RIKEN GSC) and Davis, R.W.  
(SSP/Stanford) contributed equally to this work as PIs.  
Location/Qualifiers  
1. 1570  
/organism="Arabidopsis thaliana"  
/mol\_type="mRNA"  
/db\_xref="taxon:3702"  
/chromosome="1"  
/clone="RAPL08-19-P05"  
/note="This clone is in pBluescript"  
1. 1570  
/gene="At1g74310; F1017.2"  
189. .>1570  
/gene="At1g74310; F1017.2"  
/codon\_start=1  
/product="heat shock protein 101"  
/protein\_id="AAL32674.1"  
/db\_xref="GI:17065040"  
/translation="MNPEKTHKNTETATAHAVNAGHQFTPLHAGALISDPTG  
IFPQIISAGENAAQRAERVINQALKKIPSQSPDDIPASSSLIKVIRRAQAQKS  
FGDTHLAVDQLIMGLLEDSDQIRDLNNEVATARKSEFEKLRGKGGKVVESGDTN  
RQALTKYGRDLVEQAKLDPVIGRDEIRRVIRILSRRTKNPVLIGEPGVKTAVVE  
GLAQRIVKGDVPSLTDVRLISLDMGALVAGALRGFEERLKSVLKEVEDAEGKVL  
FIDRIHLVAGAGTSGMSDANLFPMLARGQLRCIGATTLEEKYKVKVKAAPERRF  
QOYVVAEVSVPDTISILRLGKYEYEGHGVRIQDRALINAQLSARYITGSHLPDKAI  
DLVDEACANRVQSDSQFEINDLNRKMQLEIELHALEREKASKARLIEVRKELD  
DLRDKLOPLT"  
ORIGIN  
Query Match 84.0%; Score 16.8; DB 15; Length 1570;  
Best Local Similarity 90.0%; Pred. No. 8.5e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 GAAACTGACCTTCACGCCTT 20  
||||| ||||||| |||||  
Db 1458 GAAATTGAACCTTCACGCCTT 1477  
RESULT 13  
AX505540 2736 bp DNA linear PAT 27-SEP-2002  
LOCUS  
Sequence 235 from Patent WO0216655.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Arabidopsis thaliana (thale cress)  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1  
Harper, J.F., Kreps, J., Wang, X. and Zhu, T.  
Stress-regulated genes of plants, transgenic plants containing  
same, and methods of use  
Patent: WO 0216655-A 235 28-FEB-2002;  
The Scripps Research Institute (US) ; Syngenta Participations AG  
(CH)  
Location/Qualifiers  
1. 2736  
/organism="Arabidopsis thaliana"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:3702"  
ORIGIN  
Query Match 84.0%; Score 16.8; DB 6; Length 2736;  
Best Local Similarity 90.0%; Pred. No. 8.8e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 GAAACTGACCTTCACGCCTT 20  
||||| ||||||| |||||  
Db 1270 GAAATTGAACCTTCACGCCTT 1289

```

RESULT 14
AX589851
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1
Zhu,T., Glazov,E.A., Meins,F., Wang,X. and Chang,H.S.
Genes that are modulated by posttranscriptional gene silencing
Patent: WO 02081695-A 33 17-OCT-2002;
Syngenta Participations AG (CH)
FEATURES
source
1..2736
/organism="Arabidopsis thaliana"
/mol_type="unassigned DNA"
/db_xref="taxon:3702"
ORIGIN
Query Match 84.0%; Score 16.8; DB 6; Length 2736;
Best Local Similarity 90.0%; Pred. No. 8.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GAAACTGACCTTCACGCCTT 20
||||| ||||| ||||| ||||| |||||
Db 1270 GAAATTGAACCTTCACGCCTT 1289

RESULT 15
ATU13949
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 3105)
Schirmer,E.C., Lindquist,S. and Vierling,E.
An Arabidopsis heat shock protein complements a thermotolerance
defect in yeast
Plant Cell 6 (12), 1899-1909 (1994)
7866032
2 (bases 1 to 3105)
Vierling,E.
Direct Submission
Submitted (24-AUG-1994) Elizabeth Vierling, Biochemistry,
University of Arizona, Life Sciences South, Tucson, AZ 85721, USA
Location/Qualifiers
1..3105
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="1-18"
/tissue_type="leaf"
/clone_lib="lambda ZAP (Stratagene) cDNA library from heat
shock poly(A) RNA; K.W. Helm and E. Vierling, Nuc. Acids
Res., 17:7995 (1989)"
/dev_stage="adult"
/ecotype="Columbia"
1..3105
/gene="Athsp101"
<l..162
gene
5'UTR

/gene="Athsp101"
163..2898
/gene="Athsp101"
/note="this protein is a member of the highly conserved
HSP100/Clp family of proteins, and like other HSP100/Clp
proteins, Athsp101 has two consensus ATP-binding domains"
/codon_start=1
/product="Athsp101"
/db_xref="GI:537446"
/translation="MNPKEFTKNTETIATATLAHLAVNAGHAQPTPLHLAGALISDPTG
IFPQNISSAGGENAQAQSERVINGALKKLPQSPDPDIPASSSLIKVIRQAQAQKS
RGDTHLAVQIMGLLESQIRDLNVEGVATARVKSEVKLRGKRGKVSASGDTN
FOALRTYGRDLVEQKLDVPVIGRDEIRRVRIILSRRTKNPNVLIGPVGKTAIVE
GLAORTVKGDDVNSLTQVRLISLDMGALVAGAKYGEFEERLKVLEKEDEAGKVIL
FIDITHLVGAGKTEGSDAANLFPKMLARGQLRCIGATTLEEVRYKVEKDAAFERR
QQVVAEPPVDTTISILRGLKEKVEGHGVRIQDRALINAAQLSARYITGRHLPKAI
DLVDEACANVRQLOSDPEEDNLERKEMOLELILHALEREKDKAKARLLEVRKELD
DIRKQLPTWKYRKERIDEIRLKKQRELMFSLQEAERRVDLARAADLRCAIQ
EVESAIQLEGTSEENVMLTENVPHEIAEVSRWTGIPVTRLQNKERLITGLADR
LHKRVQGNQAVNAVSEAILRSRAGLGRAGQQTGSLFLGPTGVKTKLAKALBQLF
DDENLLVRIDMSEYMEQHSRLIGAPGYVGBEGQLTAVRRRPYCVILFDSEVK
AHVAVENTLQVLDGRLTDGQRTVDFRNSVIIMTSLGAHLIAGLTGKVTMEVAR
DCVMREVRKHERPELLARLDEIVDPDLSDQLRKVARLQMKDVAVRLAERGVALAVT
DAALDYILAESYDPVYGARPIRRMEKKVWTELSMNVREIDENSTVYIDGAGDLV
YRVESGGLDVASTGKSDVLIIHANGPKRSDAQAQVKQRIIEEDDDDDNEEMIED"
2899..>3105
/gene="Athsp101"
3105
/polyA_site
/gene="Athsp101"
/note="19 A nucleotides"
ORIGIN
Query Match 84.0%; Score 16.8; DB 15; Length 3105;
Best Local Similarity 90.0%; Pred. No. 8.9e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GAAACTGACCTTCACGCCTT 20
||||| ||||| ||||| ||||| |||||
Db 1432 GAAATTGAACCTTCACGCCTT 1451

RESULT 16
AY626367
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Edwardsiella ictaluri
Edwardsiella ictaluri
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Edwardsiella.
1 (bases 1 to 5094)
Thune,R.L., Fernandez,D.H., Benoit,J.L., Kelly-Smith,M.,
Rogge,M.L., Booth,N.J. and Bologna,R.A.
Identification of Virulence Factors Involved in the Pathogenesis of
Edwardsiella ictaluri Using Signature Tagged Mutagenesis
Unpublished
2 (bases 1 to 5094)
Thune,R.L., Fernandez,D.H., Benoit,J.L., Kelly-Smith,M.,
Rogge,M.L., Booth,N.J. and Bologna,R.A.
Direct Submission
Submitted (14-MAY-2004) Pathobiological Sciences, Louisiana State
University, Skip Bertman Dr. and River Rd., Baton Rouge, LA 70803,
USA
Location/Qualifiers
1..5094
/organism="Edwardsiella ictaluri"
/mol_type="genomic DNA"
/db_xref="taxon:67780"
/clone="pBK195CD"

```

CDS 56. .265  
/note="195CD-1; similar to Escherichia coli putative  
regulator for prophage"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AA141666.1"  
/db\_xref="GI:48256967"  
/translation="MGNTINTVVSAPVSYNEARLSGIPVATCGMVADGRIIRP  
KSSAQEKIEVNMIAMLKDAIANGP"  
999.1000  
misc\_feature  
/note="signature tagged mutagenesis transposon insertion  
site"  
1731. 2267  
/note="195CD-2; similar to Bordetella pertussis  
hypothetical protein"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AA141667.1"  
/db\_xref="GI:48256968"  
/translation="MAHTLIFSEALKTAILDGCKTVTRRPIKSACTRSVIKSDQWYE  
TAFKMTTLP,PERCELVIAANTPVLKASQAFMISRCPHGNAGDLIELATRADQS  
VTPPARSLKAIAIELLDITDHAALAEAGMPDTPREHFRCIWDGFYHTYPVSRWNP  
AVWLQFAPTEILAEHR"  
2556. .3266  
/note="195CD-3; putative phage protein"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AA141668.1"  
/db\_xref="GI:48256969"  
/translation="MQPCNHCLEARPIARDRAKFTALASAPLARAAGTAKNLRDPMT  
GYISRAAKSAAGIGTPTDITRAHNRAISGFVCEARPHLRIMVGRAGASTDAPGSLV  
TGSNPRVLTTPRLPELAGELSOLTREDARSQWPNVTIPRIPVSHLTPPALPHGFTV  
STPWESICRSCLTVTRCRLCVSLRYSAISQTISSPPSSAISPPARNSPCNASTRGAP  
LAPAGFIYPHRRAPHA"  
3721. .55094  
/note="195CD-4; putative replication protein for prophage"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AA141669.1"  
/db\_xref="GI:48256970"  
/translation="MSGQVPRAWGAPHPAIDPTEGRAPRTALGDAIALYAKEDKDRK  
AREQEARQRAEKQAASHQESALAAAGERFORDLPLPKIRIAHAARKDPDGPGL  
ADLMALPYLPTPLQLKLOKQEAENAGKSPASRVIRSKLALLIRIALTNS  
RFLTHYQLAVREIRDELHPLUNKKGIQTALATATASAFSGEDRQYTQPGIKQTP  
MQALAIYQLGHMALSLHTTPNWEALRPDKRKEIDLEKLPGLALSRLMCAEWTHQ  
LWRLCRWEQLEAAGQVSROTSPYLSRDALNAPRQRRLRDLFKSHELAVNEDGFT  
ILDYDVYASASPKHRRFEMWTTMKGLEIAEARGDIPFTITVTCPSRFHATTEDGH  
PNARWGSTVRDSSDLYNRPFSAPVSKKLNKLNLYWYGIKVAEPHHDGTVHWMWIFA  
HPEEREI"

ORIGIN

Query Match 84.0%; Score 16.8; DB 1; Length 5094;  
Best Local Similarity 90.0%; Pred. No. 9.1e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GAAACTGACCTTCAGCCCTT 20  
| | | | | | | | | | | | | | | | | | | | | |  
DB 3313 GCAACTGACTTTCAGCCCTT 3332

RESULT 17  
AF218796  
LOCUS AF218796 6376 bp DNA linear PLN 13-APR-2000  
DEFINITION Arabidopsis thaliana heat shock protein 101 (HSP101) gene, complete  
cds.  
ACCESSION AF218796

AF218796.1 GI:6715467  
Arabidopsis thaliana (thale cress)  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
1 (bases 1 to 6376)  
Hong, S.W. and Vierling, E.  
Mutants of Arabidopsis thaliana defective in the acquisition of  
tolerance to high temperature stress  
Proc. Natl. Acad. Sci. U.S.A. 97 (8), 4392-4397 (2000)  
10760305  
2 (bases 1 to 6376)  
Hong, S.-W. and Vierling, E.  
Direct Submission  
Submitted (22-DEC-1999) Biochemistry, University of Arizona, 1007  
E. Lowell St, Tucson, AZ 85721, USA  
FEATURES  
Location/Qualifiers  
1. 6376  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:3702"  
/chromosome="I"  
/map="ngal11-ATATPASE"  
/ecotype="Columbia"  
<1132..>4279  
/gene="HSP101"  
join(<1132..2281,2366..2547,2688..3313,3398..3875,  
3980..>4279)  
/gene="HSP101"  
/product="heat shock protein 101"  
join(1132..2281,2366..2547,2688..3313,3398..3875,  
3980..4279)  
/gene="HSP101"  
/note="clob"  
/codon\_start=1  
/product="heat shock protein 101"  
/protein\_id="AAF26423.1"  
/db\_xref="GI:6715468"

/translation="MNPEKFTHTKNTIATAHELAVNAGHAQFTPLHLAGALISDPTG  
IFPOAISAGGENAAQAEVINOALKKLPSPDPDIPASSSLIKVIRRAQAQKS  
RGDTLAVDQIMGLLEDSDQFEEDLNLERKMQLEIELHALEREKDKASARLIEVRKEL  
DLRKLQPLTMKYRKEKIDEIRLQKRELMFSLQAEARYDLARAADLYGAIQ  
EVESAIQLEGTSEENVMLENTENYRAGLGRPOQPTGSLFLGPTGVGKTELAKALAEOLF  
LHKRVQONQAVNSAAILRSRAGLGRPOQPTGSLFLGPTGVGKTELAKALAEOLF  
DDELLVRIDMSYMEQSHSRLLGAPGYVHEGGQLTAVRRPYCVILFUEVEK  
AHVAFNTLQVLDDGRITDGGRTVDFRNSVIIMTSLNGAEHLGLAGLTKVTMEVAR  
DCVNRKHFPELLNRDLDEIVVDFPLSHDQLRKVARLQMKDVAURLAERGVAVLT  
DAALDYILAESYDPVYGARPIRWMKKVVTLSKVVREEDENSTYVIDAGAGDLV  
YRVESGGIVDASTGCKSDVLIHIANGPKRSQAAQVAKQRIEIEDDDNEEMIED"

ORIGIN

Query Match 84.0%; Score 16.8; DB 15; Length 6376;  
Best Local Similarity 90.0%; Pred. No. 9.3e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GAAACTGACCTTCAGCCCTT 20  
| | | | | | | | | | | | | | | | | | | | | |  
DB 2485 GAAATTGAACCTTCAGCCCTT 2504

RESULT 18  
AC020579  
LOCUS AC020579 50821 bp DNA linear PLN 19-JAN-2001  
DEFINITION Arabidopsis thaliana chromosome 1 BAC F1017 genomic sequence,  
complete sequence.  
ACCESSION AC020579

VERSION	AC020579.5	GI:12324896	
KEYWORDS	HTG		
SOURCE	Arabidopsis thaliana (thale cress)		
ORGANISM	Arabidopsis thaliana		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eucotids II; Brassicales; Brassicaceae; Arabidopsis.		repeat_region
AUTHORS	1 (bases 1 to 50821) Lin,X., Kaul,S., Town,C.D., Benito,M.-I., Creasy,T.H., Haas,B.J., Wu,D., Maiti,R., Renning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R., Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.		repeat_region
TITLE	Arabidopsis thaliana chromosome 1 BAC F1017 genomic sequence		gene
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 50821) Lin,X. and Kaul,S.		
AUTHORS	Direct Submission		
TITLE	Submitted (05-JAN-2000) The Institute for Genomic Research, 9712		mrna
JOURNAL	Medical Center Dr. Rockville, MD 20850, USA, xlin@tigr.org		
REFERENCE	3 (bases 1 to 50821) Town,C.D. and Kaul,S.		
AUTHORS	Direct Submission		
TITLE	Submitted (19-JAN-2001) The Institute for Genomic Research, 9712		
JOURNAL	Medical Center Dr. Rockville, MD 20850, USA, cdtown@tigr.org		
COMMENT	On Jan 19, 2001 this sequence version replaced gi:12280787. Address all correspondence to:at@tigr.org		
FEATURES	BAC clone F1017 is from Arabidopsis thaliana chromosome 1 The orientation of the sequence is from SP6 to T7 end of the BAC clone. Genes were identified by a combination of several methods: Gene prediction programs including Genscan+ (Chris Burge, http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHM (Mark Borodovsky, http://genemark.biology.gatech.edu/GeneMark/), GlimmerA (a variant of GlimmerM, see Mihaela Pertea, http://www.tigr.org/softlab/glimmerm.htm/glimmerm.html, and GeneSplicer (Mihaela Pertea and Steven Salzberg, contact mpertea@tigr.org), searches of the complete sequence against a peptide database and the plant EST database at TIGR (http://www.tigr.org/tdb/tgi.shtml). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are identified by repeatmasker (Arian Smit, http://ftp.genome.washington.edu/RM/RepeatMasker.html). Location/Qualifiers 1..50821 /organism="Arabidopsis thaliana" /mol_type="genomic DNA" /db_xref="taxon:3702" /chromosomes="1" /clone="F1017" /ecotype="Columbia" /complement(9837..11757) /gene="F1017.1" /note="similar to GB:AAC49376 from [Glycine max]" complement(join(<9837..10159,10243..10343,10438..10485, 10576..10693,10800..10995,11127..11248,11404..>11757)) /gene="F1017.1" complement(join(10052..10159,10243..10343,10438..10485, 10576..10693,10800..10995,11127..11248,11404..11757)) /gene="F1017.1" /codon_start=1 /product="putative choline kinase; 11757-10052" /protein_id="AAG52400.1" /db_xref="GI:12324898" /translation="MGTEKKNVKNQYRLPREVKEALQAIASEWEDVIDSKALQVIPL KGMATNEVQIKWPREKSPKSLVRIYGGVEVFFDREDEIRTFEMSKHGHPLL LGRFNGRIIEBFHARTUSACDLRDPFISGRITATRKPFHGLEMPGAKKALLWDLRLN		
source	WTACKRLASPEEAKSFLDVMEMEINMLEKSLFDNDENIGFCHNDLQVGNIMDBEET KAITIIDYEYCYNPVAIDIANHFCEMAADYHTPHIMDYSKYPGVSEORFLKTYM SYSDKPSDTPWKKLLEDVEKYTLASHLIWGLWGTIISEHVNBIIDFYMEYARQFEQY WLTKPRLLAASEHK" complement(12470..12501) /rpt_family="AT rich" complement(13045..13067) /rpt_family="(A)n" /gene="F1017.2" /note="identical to heat shock protein 101 GB:AAF266423 from [Arabidopsis thaliana]" join(<13093..14242,14327..14508,14649..15274,15359..15836, 15941..16423) /gene="F1017.2" join(13093..14242,14327..14508,14649..15274,15359..15836, 15941..16240) /gene="F1017.2" /codon_start=1 /product="heat shock protein 101; 13093-16240" /protein_id="AAG52410.1" /db_xref="GI:12324908" /translation="WNPSEKFTHTKNTIATATHELAVNAGHAQFTPLHLAALISDPTG IFPQAISSAGSENAQAQSERVINQALKLPQSPPDIPASSSLIKVIRRAQAQKS RGDTHLAVDQLIMGLLEDSQIRDLNVEGVATARKSEVEKLRGKKGKVSASGDTN FOALKTYGRDLVQAGKLDPVIGRDEEIRRVIRILSRRTKNNPVLIGBPGVGTAVVE GLAQRIVKGDVNSLTDLRLISLDMGALVAGAKYRGEFEERLKSVLKEVEDAGKVIL FIDETHLVGAKTEGSMGAANLPKMLGQRCIGATTLEERYKYKVEDAAFRFR QOYVVAEPSVPDTISILGLAKYEGHGVIGQIPALJNAQLSARYITGRHLPKAI DLVDEACANVRQLDSPEEIDNLERKMQLEIELHALEREKDKASKARLIEVRKELD DLRKLQPLTWYRKKEKIDIRRLKQKRELMFSLQEAERRYDLARAADURYGAIO EVEGAIQLEGTSEENVMLTENYGEHIAEVSWSWTGIPVTRLGONKEKRLIGLADR LHKRVQGNQAVNAVSEAIRSRAGLGRPQPTGSGFLGPTGVGKTELAKALBQLF DDENLLVRIDMSYMEQHSVSLKIGAPGYGHEGGQLTEAVRRPYCVILFDSEVK AHAVFNTLLQVLDGRLTDCQGRTVDFPNSVIMTSLNGAEHLLAGTGTKVTMEVAR DCVREVRKHPPELLNRLDEIRVDFPLSHQOLRKVARKQMDVAVRLAERGVALAV DAALDYILAESYDPYGARIPRRWKEKKVTELSKMWVREEDENSTYVIDAGADLV YRSEGGVLDASTGKSDVLTHIANGPKRSDAAQAVKQRIEIEDDDDDNEEMIED" /gene="F1017.3" /note="similar to unknown protein GB:AAD21437 from [Arabidopsis thaliana]" complement(join(<16481..17251,17318..>17587)) /gene="F1017.3" complement(join(16481..17251,17318..17587)) /gene="F1017.3" /codon_start=1 /product="unknown protein; 17587-16481" /protein_id="AAG52411.1" /db_xref="GI:12324909" /translation="MAIGPSSVIVIGIILALTYQSILKPPPNLCSFGGPPITAPR IKLRDGRHLAYKGYLPREKAKHIVFIHSGDSRCHDAVEATLLSPDLVQRGVYVMS FDKPGYCESDPPDIPRTPKSLALDIIBELADQLSLGSKFYICKSGGQAANGCLKTYPH RLAGVTLVAPVNYIWRNLPLNISTEGNLOQKRDQWAVRAVHAPWLIYWNWTQNF PGSSVARDGGVLSPQDKDIILKLGSRKPHLAEVRQQGHIESIRNMDIVFGNWEFD PLELENPLNREGSVHLWQGDDEMLVPVTLQRYADKLPLWHLHYHEVAGGHPFLAKG VVEIVKVALNH" complement(18687..21119) /gene="F1017.4" /note="similar to unknown protein GB:AAD21437 from [Arabidopsis thaliana]" complement(join(<18687..19284,20351..20547,20799..>21119)) /gene="F1017.4" complement(join(18687..19284,20351..20547,20799..21119)) /gene="F1017.4" /codon_start=1 /product="unknown protein; 21119-18687" /protein_id="AAG52412.1" /db_xref="GI:12324910" /translation="WASDSSRSKSLKPSSRKSLFPSPVIVIGIIVALTQYSLKPP QPKLCSGSSGPPITAPRIKLDGRYLAKEHGLPREKANRKIVFIHSGDCCCHDAVFA TLLSPDLVEELGUVMSFDRPGYCESDPHPSRTPSLVSDIEEDLDQLSLGSKFYVIG KSMGGQAANGCLNKYIPHRLAGVTLVAPVNYIWRNLPLNVSTEGFNFQQRQWAV		



gap 23051..23150  
/estimated\_length=unknown  
gap 25312..25411  
/estimated\_length=unknown  
gap 29802..29901  
/estimated\_length=unknown  
gap 34331..34430  
/estimated\_length=unknown  
gap 40602..40701  
/estimated\_length=unknown  
gap 51143..51242  
/estimated\_length=unknown  
gap 65991..66090  
/estimated\_length=unknown  
gap 82664..82763  
/estimated\_length=unknown

ORIGIN

Query Match 84.0%; Score 16.8; DB 14; Length 90816;  
Best Local Similarity 90.0%; Pred. No. 1.1e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAACTGACCTTCAGCCTT 20  
||||| ||||| ||||| ||||| |||||  
DB 13227 GAACTTATCTTCAGCCTT 13208

RESULT 20  
CP000031\_40  
WPCOMMENT

Sequence split into 41 fragments LOCUS CP000031 Accession CP000031

Fragment Name	Begin	End
CP000031_00	1	110000
CP000031_01	100001	210000
CP000031_02	200001	310000
CP000031_03	300001	410000
CP000031_04	400001	510000
CP000031_05	500001	610000
CP000031_06	600001	710000
CP000031_07	700001	810000
CP000031_08	800001	910000
CP000031_09	900001	1010000
CP000031_10	1000001	1110000
CP000031_11	1100001	1210000
CP000031_12	1200001	1310000
CP000031_13	1300001	1410000
CP000031_14	1400001	1510000
CP000031_15	1500001	1610000
CP000031_16	1600001	1710000
CP000031_17	1700001	1810000
CP000031_18	1800001	1910000
CP000031_19	1900001	2010000
CP000031_20	2000001	2110000
CP000031_21	2100001	2210000
CP000031_22	2200001	2310000
CP000031_23	2300001	2410000
CP000031_24	2400001	2510000
CP000031_25	2500001	2610000
CP000031_26	2600001	2710000
CP000031_27	2700001	2810000
CP000031_28	2800001	2910000
CP000031_29	2900001	3010000
CP000031_30	3000001	3110000
CP000031_31	3100001	3210000
CP000031_32	3200001	3310000
CP000031_33	3300001	3410000
CP000031_34	3400001	3510000
CP000031_35	3500001	3610000
CP000031_36	3600001	3710000
CP000031_37	3700001	3810000
CP000031_38	3800001	3910000
CP000031_39	3900001	4010000
CP000031_40	4000001	4109442

Continuation (41 of 41) of CP000031 from base 4000001 (CP000031 Silicibacter pomeroyi DS  
Query Match 84.0%; Score 16.8; DB 1; Length 109442;  
Best Local Similarity 90.0%; Pred. No. 1.1e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAACTGACCTTCAGCCTT 20  
||||| ||||| ||||| ||||| |||||  
DB 16821 GAAACGGACCTTCGGCCTT 16840

RESULT 21  
AB128049\_24/c  
WPCOMMENT

Sequence split into 33 fragments LOCUS AB128049 Accession AB128049

Fragment Name	Begin	End
AB128049_00	1	110000
AB128049_01	100001	210000
AB128049_02	200001	310000
AB128049_03	300001	410000
AB128049_04	400001	510000
AB128049_05	500001	610000
AB128049_06	600001	710000
AB128049_07	700001	810000
AB128049_08	800001	910000
AB128049_09	900001	1010000
AB128049_10	1000001	1110000
AB128049_11	1100001	1210000
AB128049_12	1200001	1310000
AB128049_13	1300001	1410000
AB128049_14	1400001	1510000
AB128049_15	1500001	1610000
AB128049_16	1600001	1710000
AB128049_17	1700001	1810000
AB128049_18	1800001	1910000
AB128049_19	1900001	2010000
AB128049_20	2000001	2110000
AB128049_21	2100001	2210000
AB128049_22	2200001	2310000
AB128049_23	2300001	2410000
AB128049_24	2400001	2510000
AB128049_25	2500001	2610000
AB128049_26	2600001	2710000
AB128049_27	2700001	2810000
AB128049_28	2800001	2910000
AB128049_29	2900001	3010000
AB128049_30	3000001	3110000
AB128049_31	3100001	3210000
AB128049_32	3200001	3284914

Continuation (25 of 33) of AB128049 from base 2400001 (AB128049 Macaca mulatta genes, M-  
Query Match 84.0%; Score 16.8; DB 8; Length 110000;  
Best Local Similarity 90.0%; Pred. No. 1.1e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAACTGACCTTCAGCCTT 20  
||||| ||||| ||||| ||||| |||||  
DB 45479 GAACTGACCTACAGCCTT 45460

RESULT 22  
AB128049\_25/c  
WPCOMMENT

Sequence split into 33 fragments LOCUS AB128049 Accession AB128049

Fragment Name	Begin	End
AB128049_00	1	110000
AB128049_01	100001	210000
AB128049_02	200001	310000
AB128049_03	300001	410000
AB128049_04	400001	510000
AB128049_05	500001	610000
AB128049_06	600001	710000
AB128049_07	700001	810000



```

AB128049_08      800001      910000
AB128049_09      900001      1010000
AB128049_10     1000001      1110000
AB128049_11     1100001      1210000
AB128049_12     1200001      1300000
AB128049_13     1300001      1410000
AB128049_14     1400001      1510000
AB128049_15     1500001      1610000
AB128049_16     1600001      1710000
AB128049_17     1700001      1810000
AB128049_18     1800001      1910000
AB128049_19     1900001      2010000
AB128049_20     2000001      2110000
AB128049_21     2100001      2210000
AB128049_22     2200001      2310000
AB128049_23     2300001      2410000
AB128049_24     2400001      2510000
AB128049_25     2500001      2610000
AB128049_26     2600001      2710000
AB128049_27     2700001      2810000
AB128049_28     2800001      2910000
AB128049_29     2900001      3010000
AB128049_30     3000001      3110000
AB128049_31     3100001      3210000
AB128049_32     3200001      3284914
Continuation (26 of 33) of AB128049 from base 2500001 (AB128049 Macaca mulatta genes, MH

Query Match      84.0%; Score 16.8; DB 8; Length 110000;
Best Local Similarity 90.0%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1  GAAACTGACCTTCAGCCTT 20
      |||||
Db      74666  GAAACTGACCTTCAGCCTT 74647

RESULT 23
AC159748/c
LOCUS      Mus musculus 131761 bp DNA linear ROD 29-MAY-2005
DEFINITION (C57BL/6J Male) Mouse BAC Library) complete sequence.
ACCESSION AC159748.5 GI:66793523
VERSION AC159748.5
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 131761)
Muzny, D., Adams, C., Agbai II, O., Allen, C., Alsbrooks, S., Archer, P.,
Arredondo, H., Bandaranaike, D., Bangura, L., Beltran, B., Beltran, R.,
Beraducci, A., Biswal, K., Blyth, P., Bonham, H., Buhay, C., Burch, P.,
Cadore, I., Canada, A., Cardenas, V., Carter, K., Cavazos, I.,
Chacko, J., Chahrouh, M., Chavez, D., Chen, A., Chen, G., Chen, R.,
Cheng, M.-T., Chu, J., Clerck, K., Cockrell, R., Coyle, M., Cree, A.,
Curry, S., Dai, W., Davila, M.I., Davis, C., Davy-Carroll, L., De
Anda, C., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H.,
Donlin, J., McCauley, S., Dugan-Rocha, S., Dunn, A., Durbin, K.,
Dziuda, D., Egan, A., Escotto, M., Espinosa, V., Eugene, C., Fa, M.,
Fernandez, S., Fernando, P., Flagg, N., Forbes, L., Foster, P.,
Fowler, G., Fu, Q., Fuh, E., Garcia, A., Garcia, R., Garner, T.,
Gaskin, C., Gench, S., Ghose, S., Gill, R., Gonzalez, D.,
Gonzalez-Garay, M., Guevara, W., Holder, M., Haaland, W.,
Hall, B., Hamid, H., Hamilton, K., Harbes, B., Harris, R., Havlak, P.,
Hawes, A., Hawkins, E., Hayes, S., Hemphill, L., Hernandez, J.,
Hines, S., Hitchens, M., Hodgson, A., Hoques, M., Hollins, B.,
Howell, L.T., Hulyk, S., Hume, J., Imo, K., Jackson, A., Jackson, L.,
Jacob, L., Jiang, H., Johnson, B., Johnson, R., Kalafus, K., Kelly, S.,
Keys, T., Khan, Z., King, L., Kovar, C., Kowis, A., Kowis, C., Lara, F.,
Leal, S., Lee, K., Lee, S., LeGall, F.I., Lemon, S., Lewis, L., Li, B.,
Li, Y., Li, Z., Linnell, M., Liu, W., Liu, Y.-S., Liu, Y., Liyanage, D.,
London, P., Lopez, J., Lorensuhera, L., Lozano, R., Luk, T., Madu, R.,

```

Maheshwari, M., Mahoney, C., Malloy, K., Mansouri, D., Martinez, E.,  
 McClelland, H., McPherson, J., Mercadao, C., Metzker, M.,  
 Milosavljevic, A., Minja, E., Morgan, M., Morris, S., Munidasa, M.,  
 Murray, D., Nazarith, L., Ngo, D., Nguyen, N., Norwig-Eastaugh, E.,  
 Nott, A., Nwaokemele, O., Obregon, M., Ochi-Okorie, C., Odeh, E.,  
 Okwuonu, G., Okwuonu, K., Parker, D., Pasternak, S., Patel, B.,  
 Patel, V., Paul, H., Perez, A., Perez, L., Petrosino, J., Pham, T.,  
 Primus, E., Pu, L.-L., Puazo, M., Qin, X., Quinn, A., Quiroz, J.,  
 Rabata, D., Rachlin, E., Reigh, R., Ren, Y., Reuter, M., Richards, S.,  
 Rives, C., Rodriguez, F., Rojas, A., Ruiz, S.J., Sana, M., Sanders, W.,  
 Santibanez, J., Santos, R., Savary, G., Scherer, S., Shen, H., Shen, Y.,  
 Sisson, I., Sneed, A., Sodergren, E., Song, X.-Z., Sorrelle, R.,  
 Svatek, A., Taylor, E., Taylor, T., Thomas, N., Thorn, R., Thornton, R.,  
 Trejos, Z., Umani, K., Vargo, C., Verduzco, D., Villanasa, D., Virk, D.,  
 Volkov, A., Waldron, L., Walker, B., Wang, Q., Wang, S., Warren, J.,  
 Wei, X., Wheeler, D., Williams, G., Williams, R., Worley, K., Wright, R.,  
 Wu, J., Yakub, S., Yan, K., Yuan, Y., Yu, F., Zhang, J., Zhang, L.,  
 Zhang, Z., Zhou, J., Weinstock, G. and Gibbs, R.

Direct Submission  
 Unpublished  
 2 (bases 1 to 131761)  
 Worley, K.C.  
 Direct Submission  
 Submitted (09-APR-2005) Human Genome Sequencing Center, Baylor  
 College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 131761)  
 Worley, K.C.  
 Direct Submission  
 Submitted (12-MAY-2005) Human Genome Sequencing Center, Baylor  
 College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 4 (bases 1 to 131761)  
 Worley, K.C.  
 Direct Submission  
 Submitted (29-MAY-2005) Human Genome Sequencing Center, Baylor  
 College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 On May 29, 2005 this sequence version replaced gi:63253324.  
 Sequencing is completed to a minimum standard of double strand  
 coverage with a minimum of 2 clones and 2 reads with no ambiguities  
 or 2 chemistries with a minimum of 2 clones and 3 reads with no  
 ambiguities. If the sequence quality does not meet this standard,  
 it will be indicated in the annotation.

The repeat regions shown were identified using RepeatMasker by  
 Adrian Smit.

Sequence similarities were identified using Powerblast by Jinghui  
 Zhang.

Exon/Intron boundaries of identified genes were chosen if there  
 were canonical splice junctions that maintained sequence continuity  
 across the splice junctions.

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Drafting Center Code: WIBR  
 Contact: hgsc-help@bcm.tmc.edu.

FEATURES  
 source

1. .131761  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="10"  
 /clone="RP24-363022"  
 complement(1. .14781)  
 /note="overlaps bases 1. .14781 of clone AC157020"  
 /function="clone overlap"

repeat\_region

952. 1027  
 /rpt\_family="(TG)n"

repeat\_region

1027. 1090  
 /rpt\_family="(TC)n"

repeat\_region

1091. 1113  
 /rpt\_family="(TG)n"





The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/projects/c\\_elegans/wormpep](http://www.sanger.ac.uk/projects/c_elegans/wormpep) RP23-256C7 is from the RPCI-23 Mouse BAC Library constructed by the group of Pieter de Jong.

For further details see <http://www.chori.org/bacpac/home.htm>

# VECTOR: pBACE3.6.

## Location/Qualifiers

```
1. /length=158009
   /organism="Mus musculus"
   /mol_type="genomic DNA"
   /db_xref="taxon:10090"
   /chromosome="4"
   /clone="RP23-256C7"
   /clone_lib="RPCI-23"
```

## ORIGIN

```
Query Match      84.0%; Score 16.8; DB 9; Length 158009;
Best Local Similarity 90.0%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 GAAACTGACCTTCAGCCCTT 20
|||||  |||||  |||||  |||||
Db 150633 GAAACTGACCTTCACCCCTT 150652
```

## RESULT 25

### AC160257

#### LOCUS

```
AC160257 166168 bp DNA linear HTG 01-JUL-2005
Bos taurus clone CH240-92E3, *** SEQUENCING IN PROGRESS ***, 20
unordered pieces.
```

#### AC160257

```
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
```

#### KEYWORDS

```
Bos taurus (cow)
```

#### SOURCE

#### ORGANISM

```
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Bovinae; Bos.
```

#### REFERENCE

#### AUTHORS

1 (bases 1 to 166168)  
Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, X., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, N., Hoques, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, S., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuewa, L., Loulsegged, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, W., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., McLeod, M. P., McNeill, T. Z., Meenen, E., Mahoney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,

Nwaokemele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkoch, C., Popper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., L., Puzo, M., Quiros, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D. von Niederhauser, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

## TITLE

### JOURNAL

### REFERENCE

### AUTHORS

### TITLE

### JOURNAL

### REFERENCE

### AUTHORS

### TITLE

### JOURNAL

### REFERENCE

### AUTHORS

### TITLE

### JOURNAL

### REFERENCE

### AUTHORS

### TITLE

### JOURNAL

### REFERENCE

### AUTHORS

### TITLE

### JOURNAL

### REFERENCE

### AUTHORS

### TITLE

### JOURNAL

### REFERENCE

### AUTHORS

### TITLE

### JOURNAL

### REFERENCE

### AUTHORS

### TITLE

### JOURNAL

### REFERENCE

### AUTHORS

### TITLE

### JOURNAL

### REFERENCE

### AUTHORS

### TITLE

### JOURNAL

### REFERENCE

### AUTHORS

### TITLE

### JOURNAL

### REFERENCE

### AUTHORS

### TITLE

### JOURNAL

### REFERENCE

### AUTHORS

### TITLE

### JOURNAL

### REFERENCE

### AUTHORS

### TITLE

### JOURNAL

### REFERENCE

### AUTHORS

### TITLE

### JOURNAL

### REFERENCE

### AUTHORS

### TITLE

### JOURNAL

### REFERENCE

### AUTHORS

Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu

Center project name: PFNE

Center clone name: CH240-92E3

Consensus quality: 160105 bases at least Q40

Consensus quality: 161606 bases at least Q30

Consensus quality: 162879 bases at least Q20

Estimated insert size: 165468; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

Consensus quality: 160105 bases at least Q40

Consensus quality: 161606 bases at least Q30

Consensus quality: 162879 bases at least Q20

Estimated insert size: 165468; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

Consensus quality: 160105 bases at least Q40

Consensus quality: 161606 bases at least Q30

Consensus quality: 162879 bases at least Q20

Estimated insert size: 165468; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

Consensus quality: 160105 bases at least Q40

Consensus quality: 161606 bases at least Q30

Consensus quality: 162879 bases at least Q20

Estimated insert size: 165468; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

Consensus quality: 160105 bases at least Q40

Consensus quality: 161606 bases at least Q30

Consensus quality: 162879 bases at least Q20

Estimated insert size: 165468; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

Consensus quality: 160105 bases at least Q40

Consensus quality: 161606 bases at least Q30

Consensus quality: 162879 bases at least Q20

Estimated insert size: 165468; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

Consensus quality: 160105 bases at least Q40

Consensus quality: 161606 bases at least Q30

Consensus quality: 162879 bases at least Q20

Estimated insert size: 165468; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently  
consists of 20 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

\* 1 2274: contig of 2274 bp in length  
2275 2324: gap of 50 bp

\* 2325 5443: contig of 3119 bp in length  
\* 5444 5493: gap of 50 bp  
\* 5494 25127: contig of 19634 bp in length  
\* 25128 25177: gap of 50 bp  
\* 25178 29619: contig of 4442 bp in length  
\* 29620 29669: gap of 50 bp  
\* 29670 34051: contig of 4382 bp in length  
\* 34052 34101: gap of 50 bp  
\* 34102 38215: contig of 4114 bp in length  
\* 38216 38301: gap of 86 bp  
\* 38302 50698: contig of 12397 bp in length  
\* 50699 50748: gap of 50 bp  
\* 50749 80138: contig of 29390 bp in length  
\* 80139 80188: gap of 50 bp  
\* 80189 86672: contig of 6484 bp in length  
\* 86673 86722: gap of 50 bp  
\* 86723 89421: contig of 2699 bp in length  
\* 89422 89471: gap of 50 bp  
\* 89472 93059: contig of 3588 bp in length  
\* 93060 93397: gap of 338 bp  
\* 93398 94145: contig of 748 bp in length  
\* 94146 94195: gap of 50 bp  
\* 94196 116317: contig of 22122 bp in length  
\* 116318 116367: gap of 50 bp  
\* 116368 122973: contig of 6606 bp in length  
\* 122974 123023: gap of 50 bp  
\* 123024 147038: contig of 24015 bp in length  
\* 147039 147088: gap of 50 bp  
\* 147089 158106: contig of 11018 bp in length  
\* 158107 158156: gap of 50 bp  
\* 158157 160798: contig of 2642 bp in length  
\* 160799 160898: gap of unknown length  
\* 160899 161903: contig of 1005 bp in length  
\* 161904 162003: gap of unknown length  
\* 162004 163414: contig of 1411 bp in length  
\* 163415 163514: gap of unknown length  
\* 163515 166168: contig of 2654 bp in length.

Location/Qualifiers  
1. .166168  
/organism="Bos taurus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9913"  
/clone="CH240-92E3"  
2275. .2324  
/estimated\_length=50  
5444. .5493  
/estimated\_length=50  
25128. .25177  
/estimated\_length=50  
29620. .29669  
/estimated\_length=50  
34052. .34101  
/estimated\_length=50  
38216. .38301  
/estimated\_length=86  
50699. .50748  
/estimated\_length=50  
80139. .80188  
/estimated\_length=50  
86673. .86722  
/estimated\_length=50  
89422. .89471  
/estimated\_length=50  
93060. .93397  
/estimated\_length=338  
94146. .94195  
/estimated\_length=50  
116318. .116367  
/estimated\_length=50  
122974. .123023  
/estimated\_length=50  
147039. .147088  
/estimated\_length=50

FEATURES  
source

gap 158107. .158156  
/estimated\_length=50  
gap 160799. .160898  
/estimated\_length=unknown  
gap 161904. .162003  
/estimated\_length=unknown  
gap 163415. .163514  
/estimated\_length=unknown

ORIGIN  
Query Match 84.0%; Score 16.8; DB 14; Length 166168;  
Best Local Similarity 90.0%; Pred.No. 1.1e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GAAACTGACCTTCAGCCTT 20  
||||| ||||| ||||| |||||  
Db 22623 GAAACAGACCTTACGCCTT 22642  
||||| ||||| ||||| |||||

RESULT 26  
AC123803  
LOCUS 169964 bp DNA linear ROD 27-NOV-2003  
DEFINITION Mus musculus BAC clone RP24-492115 from chromosome 10, complete  
sequence.  
AC123803  
ACCESSION AC123803.4 GI:33457231  
VERSION HTG.  
KEYWORDS Mus musculus (house mouse)  
SOURCE Mus musculus  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidea; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 169964)  
AUTHORS Griffin,M. and Haglund,K.  
TITLE The sequence of Mus musculus BAC clone RP24-492115  
JOURNAL Unpublished (2001)  
REFERENCE 2 (bases 1 to 169964)  
AUTHORS Wilson,R.  
TITLE Sequencing of Mus musculus  
JOURNAL Unpublished (2001)  
REFERENCE 3 (bases 1 to 169964)  
AUTHORS McPherson,J.D. and Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (01-JUN-2002) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
REFERENCE 4 (bases 1 to 169964)  
AUTHORS McPherson,J.D. and Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (08-MAY-2003) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
REFERENCE 5 (bases 1 to 169964)  
AUTHORS Wilson,R.K.  
TITLE Direct Submission  
JOURNAL Submitted (06-AUG-2003) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
REFERENCE 6 (bases 1 to 169964)  
AUTHORS Wilson,R.  
TITLE Direct Submission  
JOURNAL Submitted (27-NOV-2003) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Aug 6, 2003 this sequence version replaced gi:30425596.  
----- Genome Center  
Center= Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu>  
Contact: [submissions@wustl.edu](mailto:submissions@wustl.edu)  
----- Summary Statistics  
-----  
Center project name: M\_BB0492115  
-----

NOTICE: This sequence may not represent the entire insert of this  
clone. It may be shorter because we only sequence overlapping  
clone sections once, or longer because we provide a small overlap

between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

#### SOURCE INFORMATION:

The RPCI-24 BAC Library has been constructed by Pieter de Jong and coworkers (<http://www.chori.org>) from male C57BL/6J mouse spleen and/or brain genomic DNA. The clone and detailed information can be obtained from Pieter de Jong and coworkers at <http://www.chori.org>

#### NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone. This clone is overlapped by AC123944.

#### FEATURES

source	Location/Qualifiers
--------	---------------------

1..169964	/organism="Mus musculus"
	/mol_type="genomic DNA"
	/db_xref="taxon:10090"
	/chromosome="10"
	/map="10"
	/clone="RP24-492115"
	/clone_lib="RPCI-24"

repeat_region	1512..1704
---------------	------------

repeat_region	/rpt_family="MIR"
---------------	-------------------

repeat_region	1819..1926
---------------	------------

repeat_region	/rpt_family="Alu"
---------------	-------------------

repeat_region	2509..4911
---------------	------------

repeat_region	/rpt_family="L1"
---------------	------------------

repeat_region	4948..5232
---------------	------------

repeat_region	/rpt_family="L1"
---------------	------------------

repeat_region	6413..6522
---------------	------------

repeat_region	/rpt_family="MIR"
---------------	-------------------

repeat_region	8658..8733
---------------	------------

repeat_region	/rpt_family="ID"
---------------	------------------

repeat_region	9049..9385
---------------	------------

repeat_region	/rpt_family="L1"
---------------	------------------

repeat_region	9385..9635
---------------	------------

repeat_region	/rpt_family="L1"
---------------	------------------

repeat_region	9692..10027
---------------	-------------

repeat_region	/rpt_family="L1"
---------------	------------------

repeat_region	10041..10119
---------------	--------------

repeat_region	/rpt_family="L1"
---------------	------------------

repeat_region	10184..10720
---------------	--------------

repeat_region	/rpt_family="L1"
---------------	------------------

repeat_region	10723..12473
---------------	--------------

repeat_region	/rpt_family="L1"
---------------	------------------

repeat_region	12619..12885
---------------	--------------

repeat_region	/rpt_family="B4"
---------------	------------------

repeat_region	13038..13154
---------------	--------------

repeat_region	/rpt_family="Alu"
---------------	-------------------

repeat_region	13356..13531
---------------	--------------

repeat_region	/rpt_family="L1"
---------------	------------------

repeat_region	13946..14260
---------------	--------------

repeat_region	/rpt_family="L1"
---------------	------------------

repeat_region	14261..14642
---------------	--------------

repeat_region	/rpt_family="MaLR"
---------------	--------------------

repeat_region	14643..15014
---------------	--------------

repeat_region	/rpt_family="L1"
---------------	------------------

repeat_region	15040..15231
---------------	--------------

repeat_region	/rpt_family="B4"
---------------	------------------

repeat_region	15232..15273
---------------	--------------

repeat_region	/rpt_family="B2"
repeat_region	15274..15320
repeat_region	/rpt_family="L1"
repeat_region	15451..15628
repeat_region	/rpt_family="MaLR"
repeat_region	18196..18990
repeat_region	/rpt_family="L1"
repeat_region	19177..19629
repeat_region	/rpt_family="L1"
repeat_region	19992..20131
repeat_region	/rpt_family="Alu"
repeat_region	20583..20683
repeat_region	/rpt_family="Alu"
repeat_region	20735..20795
repeat_region	/rpt_family="ERV1"
repeat_region	21177..21241
repeat_region	/rpt_family="ERV1"
repeat_region	21224..21359
repeat_region	/rpt_family="ERV1"
repeat_region	21427..21605
repeat_region	/rpt_family="ERV1"
repeat_region	22190..22414
repeat_region	/rpt_family="RMER6A"
repeat_region	22430..22624
repeat_region	/rpt_family="L1"
repeat_region	22619..23179
repeat_region	/rpt_family="L1"
repeat_region	23180..23580
repeat_region	/rpt_family="RMER6A"
repeat_region	25016..25423
repeat_region	/rpt_family="MaLR"
repeat_region	26531..26586
repeat_region	/rpt_family="ERV1"
repeat_region	28191..28334
repeat_region	/rpt_family="ERVK"
repeat_region	28335..28737
repeat_region	/rpt_family="MaLR"
repeat_region	28738..28901
repeat_region	/rpt_family="ERVK"
repeat_region	29230..29295
repeat_region	/rpt_family="CR1"
repeat_region	29654..29766
repeat_region	/rpt_family="L2"
repeat_region	29890..30166
repeat_region	/rpt_family="MER2_type"
repeat_region	30184..30275
repeat_region	/rpt_family="L2"
repeat_region	30394..30571
repeat_region	/rpt_family="L1"
repeat_region	30768..30941
repeat_region	/rpt_family="L1"
repeat_region	31083..31139
repeat_region	/rpt_family="L1"
repeat_region	31140..31533
repeat_region	/rpt_family="MaLR"
repeat_region	31534..31838
repeat_region	/rpt_family="L1"
repeat_region	31891..32087
repeat_region	/rpt_family="MIR"
repeat_region	32408..32471
repeat_region	/rpt_family="tRNA-Ala-GCA"
repeat_region	32512..32573
repeat_region	/rpt_family="ERV"
repeat_region	32574..32678
repeat_region	/rpt_family="Alu"
repeat_region	32679..32783
repeat_region	/rpt_family="ERV"
repeat_region	32889..33364
repeat_region	/rpt_family="ERVK"
repeat_region	33652..33782
repeat_region	/rpt_family="MaLR"
repeat_region	34129..34280
repeat_region	/rpt_family="MaLR"



```

/estimated_length=unknown
72905..73004
/estimated_length=unknown
80226..80325
/estimated_length=unknown
87033..87132
/estimated_length=unknown
91606..91705
/estimated_length=unknown
93135..93234
/estimated_length=unknown
102020..102119
/estimated_length=unknown
107692..107791
/estimated_length=unknown
111470..111569
/estimated_length=unknown
113108..113207
/estimated_length=unknown
139895..139994
/estimated_length=unknown
143526..143625
/estimated_length=unknown
144688..144787
/estimated_length=unknown
147006..147105
/estimated_length=unknown
166370..166469
/estimated_length=unknown

```

```

Query Match      84.0%; Score 16.8; DB 14; Length 170054;
Best Local Similarity 90.0%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      1  GAAACTGACCTTCAGCGCTT 20
        |||||  |||||  |||||  |||||
Db      9193 GAAACTGACTTTCGCGCTT 9212

```

```

RESULT 28
AC160981/c
LOCUS      AC160981      175606 bp      DNA      linear      ROD 01-JUL-2005
DEFINITION Mus musculus BAC clone RP23-363D11 from chromosome 9, complete
sequence.
ACCESSION  AC160981 AC126423
VERSION    AC160981.2 GI:68448658
KEYWORDS   HTG.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
           Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 175606)
AUTHORS   Wang, C., Lewis, S., Cotton, M. and Levy, A.
JOURNAL   The sequence of Mus musculus BAC clone RP23-363D11
TITLE     Unpublished (2001)
REFERENCE  2 (bases 1 to 175606)
AUTHORS   Wilson, R.K.
JOURNAL   Direct Submission
TITLE     Submitted (05-MAY-2005) Genome Sequencing Center, 4444 Forest Park
         Parkway, St. Louis, MO 63108, USA
REFERENCE  3 (bases 1 to 175606)
AUTHORS   Wilson, R.K.
JOURNAL   Direct Submission
TITLE     Submitted (01-JUL-2005) Genome Sequencing Center, Washington
         University School of Medicine, 4444 Forest Park Parkway, St. Louis,
         MO 63108, USA
COMMENT   On Jul 1, 2005 this sequence version replaced gi:63028318.
         ----- Genome Center
         Center: Washington University Genome Sequencing Center
         Center code: WUGSC
         Web site: http://genome.wustl.edu

```

```

Contact: submissions@watson.wustl.edu
----- Summary Statistics
Center project name: M_BA0363D11
Drafting center: WIBR

```

## NOTICE:

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e. phred quality >=30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone, fosmid clone or direct clone walk sequence. Sequence from the Mouse Genome Sequencing Consortium whole genome shotgun may have been used to obtain the consensus sequence. The assembly was confirmed by restriction digest.

This finishing standard has slightly changed from the previous Human standard. Specifically, standards for regions of low sequence complexity (such as dinucleotide repeats and small unit tandem repeats) have been relaxed. These regions are very prevalent in the mouse genome, and the return on extended finishing efforts is minimal.

If a sequence meets the criteria of the above statement, it needs no comments or tags. If the criteria are not met, such as ambiguous bases, then the region is duly annotated.

## MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

## SOURCE INFORMATION:

The BAC library has been constructed by Kazutoyo Oseawa and Minako Tatenio in the laboratory of Pieter de Jong (<http://www.chori.org>) from female C57BL/6J mouse kidney and/or brain genomic DNA. The clone and detailed information can be obtained from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

This sequence is the entire insert of the clone.

```

FEATURES
source
1. .175606
   /organism="Mus musculus"
   /mol_type="genomic DNA"
   /db_xref="taxon:10090"
   /chromosome="9"
   /clone="RP23-363D11"
   /clone_lib="RPCI-23"
   54333..54402
   /note="Sequence derived from one plasmid subclone."
   55808..55931
   /note="Sequence derived from one plasmid subclone."
   56262..56342
   /note="Sequence derived from one plasmid subclone."
   57795..57880
   /note="Sequence derived from one plasmid subclone."
   58817..58883
   /note="Sequence derived from PCR product of project DNA"
   62541..62544
   /note="Sequence derived from one plasmid subclone."
   74086..75848
   /note="Unresolved tandem repeat."
   162778..162891
   /note="Unresolved simple sequence repeat."
   162892..162905
   /note="Unresolved homopolymeric repeat."

```

## ORIGIN

```

Query Match      84.0%; Score 16.8; DB 9; Length 175606;
Best Local Similarity 90.0%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Oy 1 GAAACTGACCTTCACGCCTT 20
| | | | | | | | | | | | | | | | | |
Db 15158 GCATCTGACCTTCACGCCTT 15139

RESULT 29
AC025795 182813 bp DNA linear HTG 21-MAR-2003
LOCUS Mus musculus chromosome 4 clone RP23-115E21 map 4, WORKING DRAFT
DEFINITION
ACCESSION AC025795
VERSION AC025795.5 GI:29135649
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
1 (bases 1 to 182813)
Mus musculus chromosome 4, clone RP23-115E21
Unpublished
2 (bases 1 to 182813)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Balwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campoliano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczyk,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Melidrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (14-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 182813)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cook,A., Cooke,P., Corum,B., DeArelano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., Maclean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Melidrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,W.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (21-MAR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 21, 2003 this sequence version replaced gi:13357484.
All repeats were identified using RepeatMasker:

```

```

Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WtBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8060
Center clone name: 115_E_21
----- Summary Statistics
Sequencing vector: M13; M77815; 32k of reads
Sequencing method: Plasmid; n/a; 68k of reads
Chemistry: Dye-terminator Big Dye; 100k of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 182781 bases at least Q40
Consensus quality: 182811 bases at least Q30
Consensus quality: 182813 bases at least Q20
Insert size: 170000; agarose-fp
Insert size: 182813; sum-of-fragments
Quality coverage: 14.0 in Q20 bases; agarose-fp
Quality coverage: 13.0 in Q20.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 182813: contig of 182813 bp in length.
assembly fragment.
Location/Qualifiers
1..182813
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="4"
/map="4"
/clones="RP23-115E21"
/clone_lib="RPCI-23 Female Mouse BAC"

ORIGIN
Query Match 84.0%; Score 16.8; DB 14; Length 182813;
Best Local Similarity 90.0%; Pred. No. 1.2e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 GAAACTGACCTTCACGCCTT 20
| | | | | | | | | | | | | | | | | |
Db 28536 GAAACTGACCTTCACGCCTT 28555

RESULT 30
AC150217 187385 bp DNA linear HTG 29-JUL-2004
LOCUS Callithrix jacchus clone CH259-250C3, WORKING DRAFT SEQUENCE, 5
DEFINITION
ACCESSION AC150217
VERSION AC150217.2 GI:50811676
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE Callithrix jacchus (white-tufted-ear marmoset)
ORGANISM Callithrix jacchus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Platyrrhini;
Callitrichidae; Callithrix.
1 (bases 1 to 187385)
Antonelias,A., Ayele,K., Benjamin,B., Blakesley,R.W.,
Bouffard,G.G., Brinkley,C., Brooks,S., Chu,G., Coleman,B.,
Coleman,H., Daki,N., Engle,J., Guan,X., Gupta,J., Haghighi,P.,
Han,J., Hansen,N., Ho,S.-L., Hu,P., Hurlie,B., Idol,J.R., Jones,C.,
Karlsne,E., Kim,H., Kwong,P., Laric,P., Larson,S., Lee-Lin,S.-Q.,
Legaspi,R., Madden,M., Maduro,Q.L., Maduro,V.B., Margulies,E.H.,

```



Masiello,C., Maskeri,B., McDowell,J., Mullikin,J.C., Paguirigan,C., Park,M., Portnoy,M.E., Prasad,A., Puri,O., Reddix-Dugue,N., Schandler,K., Schueler,M.G., Shah,K., Sison,C., Staniripop,S., Thomas,J.W., Thomas,P.J., Tsipouri,V., Vogt,J.L., Wetherby,K.D., Young,A. and Green,E.D.

# TITLE JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

2 (bases 1 to 187385)

Green,E.D.

Direct Submission

Submitted (08-JUL-2004) NIH Intramural Sequencing Center, 8717

Grovmont Circle, Gaithersburg, MD 20877, USA

3 (bases 1 to 187385)

Green,E.D.

Direct Submission

Submitted (29-JUL-2004) NIH Intramural Sequencing Center, 8717

Grovmont Circle, Gaithersburg, MD 20877, USA

On Jul 29, 2004 this sequence version replaced gi:50057861.

## ----- Genome Center

Center: NIH Intramural Sequencing Center

Center code: NISC

Web site: <http://www.nisc.nih.gov>

Contact: [nisc\\_zoo@nhgri.nih.gov](mailto:nisc_zoo@nhgri.nih.gov)

----- Project Information

Center project name: fra

Center clone name: 250C03

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8x average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 186482 bases at least Q40

Consensus quality: 186784 bases at least Q30

Consensus quality: 186922 bases at least Q20

Insert size: 184000; agarose-fp

Insert size: 186985; sum-of-contigs

Quality coverage: 9.97x in Q20 bases; agarose-fp

Quality coverage: 9.81x in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently consists of 5 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

\* This sequence will be replaced

\* by the finished sequence as soon as it is available and

\* the accession number will be preserved.

\* 1 21558: contig of 21558 bp in length

\* 21559 21658: gap of unknown length

\* 21659 102567: contig of 80909 bp in length

\* 102568 102667: gap of unknown length

\* 102668 114064: contig of 11397 bp in length

\* 114065 114164: gap of unknown length

\* 114165 139582: contig of 25418 bp in length

\* 139583 139582: gap of unknown length

\* 139583 187385: contig of 47703 bp in length.

Location/Qualifiers

1. .187385

/organism="Callithrix jacchus"

/mol\_type="genomic DNA"

## FEATURES

## source

/db\_xref="taxon:9483"  
/clone="CH259-250C3"  
/clone\_lib="CH259"  
/note="TBAC resource: <http://bacpac.chori.org/>"  
1. .26014  
/note="clone overlaps with GenBank Accession Number  
AC150008 clone CH259-154D8 (center project name fgz)"  
misc\_feature  
1. .21558  
/note="assembly\_fragment  
clone\_end:SP6  
vector\_side:left"  
21559. .21658  
/estimated length=unknown  
21659. .102567  
/note="assembly\_fragment"  
102568. .102667  
/estimated length=unknown  
102668. .114064  
/note="assembly\_fragment"  
114065. .114164  
/estimated length=unknown  
114165. .139582  
/note="assembly\_fragment"  
139583. .139682  
/estimated length=unknown  
139683. .187385  
/note="assembly\_fragment  
clone\_end:T7  
vector\_side:right"  
ORIGIN

Query Match 84.0%; Score 16.8; DB 14; Length 187385;  
Best Local Similarity 90.0%; Pred. No. 1.2e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GAAACTGACCTTCACGCCCTT 20

|||||

Db 133817 GAAACTGACCTTCACGCCCTT 133836

## RESULT 31

AC119772/c

LOCUS

DEFINITION

AC119772

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Rattus norvegicus

HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.

HTG 20-NOV-2002

WORKING DRAFT SEQUENCES, 4

unordered pieces.

1 (bases 1 to 187484)

REFERENCE

AUTHORS

Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,P., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dedeich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Evans,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,B., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,M., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,



Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, H., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Loulseghe, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K., Nwaokeme, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Prannkoc, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, B., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstein, G. and Gibbs, R.A.

## TITLE

Direct Submission

## JOURNAL

Unpublished

2 (bases 1 to 187484)

## REFERENCE

Worley, K.C.

Direct Submission

## JOURNAL

Submitted (02-MAY-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 187484)

## REFERENCE

Rat Genome Sequencing Consortium.

## JOURNAL

Submitted (20-NOV-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Nov 20, 2002 this sequence version replaced gi:23808843.

The sequence in this assembly is a combination of BAC based reads

and whole genome shotgun sequencing reads assembled using Atlas

(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described

in the feature table below represents a scaffold in the Atlas

assembly (a 'contig-scaffold'). Within each contig-scaffold,

individual sequence contigs are ordered and oriented, and separated

by sized gaps filled with Ns to the estimated size. The sequence

may extend beyond the ends of the clone and there may be sequence

contigs within a contig-scaffold that consist entirely of whole

genome shotgun sequence reads. Both end sequences and whole genome

shotgun sequence only contigs will be indicated in the feature

table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GVOV

Center clone name: CH230-390M12

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 177051 bases at least Q40

Consensus quality: 179151 bases at least Q30

Consensus quality: 180782 bases at least Q20

Estimated insert size: 181642; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length  
 \* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 4 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 147431: contig of 147431 bp in length  
 \* 147432 147531: gap of unknown length  
 \* 147532 185118: contig of 37587 bp in length  
 \* 185119 185218: gap of unknown length  
 \* 185219 186303: contig of 1085 bp in length  
 \* 186304 186403: gap of unknown length  
 \* 186404 187484: contig of 1081 bp in length.

## FEATURES

## source

1..187484  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-390M12"

## misc\_feature

1..1114  
 /note="wgs\_end\_extension  
 clone\_end:Sp6"  
 2506..3438  
 /note="clone\_boundary  
 clone\_end:Sp6  
 site:

## misc\_feature

end\_sequence:BZ204516"  
 complement(30054..32919)  
 /note="clone\_boundary  
 clone\_end:T7  
 site:

## gap

end\_sequence:BZ204514"  
 147432..147531

## misc\_feature

/estimated\_length=unknown  
 181683..183091  
 /note="wgs\_end\_extension  
 clone\_end:T7"  
 183586..185118

## misc\_feature

/note="wgs\_end\_extension  
 clone\_end:T7"  
 185119..185218  
 /estimated\_length=unknown  
 186304..186403

## gap

/estimated\_length=unknown

## ORIGIN

Query Match 84.0%; Score 16.8; DB 14; Length 187484;  
 Best Local Similarity 90.0%; Pred. No. 1.2e+03;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACGCCTT 20

|||||  
 Db 132581 GAAACTGACCTTCACGCCTT 132562

## RESULT 32

AC148670

LOCUS

AC148670 189655 bp DNA linear PRI 12-APR-2004  
 Macaca mulatta Major Histocompatibility Complex BAC MMU063G23,  
 complete sequence.

## ACCESSION

AC148670

VERSION

HTG.

## KEYWORDS

Macaca mulatta (rhesus monkey)

## SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Cercopithecoidea; Cercopithecinae; Macaca.

REFERENCE

1 (bases 1 to 189655)

AUTHORS Daza-Vamanta,R., Guthrie,B., Glusman,G., Rowan,L., Vu,Q.,  
 TITLE Williams,L.M., Kurihara,B.K. and Geraghty,D.E.  
 Immunological Divergence Between Rhesus Monkeys (Macaca mulatta)  
 and Human as Revealed by the Complete Sequence of a Macaque Major  
 Histocompatibility Complex (MHC)  
 JOURNAL Unpublished  
 REMARK The BAC is one of 59 projects that collectively comprise a rhesus  
 macaque major histocompatibility complex genomic sequence.  
 REFERENCE 2 (bases 1 to 189655)  
 AUTHORS Daza-Vamanta,R., Guthrie,B., Glusman,G., Rowan,L., Vu,Q.,  
 TITLE Williams,L.M., Kurihara,B.K. and Geraghty,D.E.  
 JOURNAL Direct Submission  
 JOURNAL Submitted (12-APR-2004) Clinical Research, Fred Hutchinson Cancer  
 Research Center, 1100 Fairview Avenue North, Seattle, WA 98109, USA  
 FEATURES  
 source  
 1..189655  
 /organism="Macaca mulatta"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9544"  
 /clone="BAC MMU063G23"  
 /clone\_lib="The CHORI-250 Rhesus macaque BAC library  
 constructed at the Children's Hospital Oakland Research  
 Institute, BACPAC Resources"

ORIGIN  
 Query Match 84.0%; Score 16.8; DB 8; Length 189655;  
 Best Local Similarity 90.0%; Pred. No. 1.2e+03;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCAGCCTT 20  
 ||||| ||||| |||||  
 DB 183623 GAAACTGACCTTCAGCCTT 183642

RESULT 33  
 AC146618 197233 bp DNA linear ROD 06-FEB-2004  
 LOCUS Mus musculus BAC clone RP23-5P3 from chromosome 10, complete  
 DEFINITION sequence.  
 AC146618  
 AC146618.2 GI:41351755  
 VERSION HTG.  
 KEYWORDS Mus musculus (house mouse)  
 SOURCE Mus musculus  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.  
 1 (bases 1 to 197233)  
 Haakenson,W., Kozlowski,A. and Haglund,K.  
 REFERENCE 1 The sequence of Mus musculus BAC clone RP23-5P3  
 AUTHORS Unpublished (2001)  
 JOURNAL Unpublished (2001)  
 REFERENCE 2 (bases 1 to 197233)  
 AUTHORS Wilson,R.  
 TITLE Sequencing of Mus musculus  
 JOURNAL Unpublished (2001)  
 REFERENCE 3 (bases 1 to 197233)  
 AUTHORS Wilson,R.K.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-SEP-2003) Genome Sequencing Center, 4444 Forest Park  
 Parkway, St. Louis, MO 63108, USA  
 4 (bases 1 to 197233)  
 Wilson,R.K.  
 TITLE Direct Submission  
 JOURNAL Submitted (28-JAN-2004) Genome Sequencing Center, 4444 Forest Park  
 Parkway, St. Louis, MO 63108, USA  
 5 (bases 1 to 197233)  
 Wilson,R.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-FEB-2004) Department of Genetics, Washington  
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 On Jan 28, 2004 this sequence version replaced gi:34495113.  
 ----- Genome Center

Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu>  
 Contact: [submissions@wustl.edu](mailto:submissions@wustl.edu)  
 ----- Summary Statistics  
 -----  
 Center project name: M\_BA0005P03  
 -----

NOTICE: This sequence may not represent the entire insert of this  
 clone. It may be shorter because we only sequence overlapping  
 clone sections once, or longer because we provide a small overlap  
 between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate  
 chemistry, or covered by high quality data (i.e., phred quality >=  
 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone, fosmid clone or direct clone walk sequence.  
 Sequence from the Mouse Genome Sequencing Consortium whole genome  
 shotgun may have been used to obtain the consensus sequence; and  
 the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
 Mapping information for this clone was provided by Dr. Wes Warren,  
 Department of Genetics, Washington University, St. Louis MO. For  
 additional information about the map position of this sequence, see  
<http://genome.wustl.edu>

SOURCE INFORMATION:  
 The RPCI-23 BAC Library has been constructed by Kazutoyo Oseawa  
 and Minako Tatenio in the laboratory of Pieter de Jong  
 (<http://www.chori.org>) from female C57BL/6J mouse kidney and/or  
 brain genomic DNA. The clone and detailed information can be  
 obtained from Research Genetics, Inc. (<http://www.resgen.com>) or  
 Pieter de Jong and coworkers at <http://www.chori.org>

NEIGHBORING SEQUENCE INFORMATION:  
 This sequence is the entire insert of the clone. This clone is  
 overlapped by AC123803.

FEATURES	source
Location/Qualifiers	1..197233
organism="Mus musculus"	
mol_type="genomic DNA"	
db_xref="taxon:10090"	
chromosome="10"	
map="10"	
clone="RP23-5P3"	
clone_lib="RPCI-23"	
repeat_region	3..184
rpt_family="ERV1"	
repeat_region	238..412
rpt_family="ERV1"	
repeat_region	427..974
rpt_family="L1"	
repeat_region	998..1192
rpt_family="ERV1"	
repeat_region	1195..1402
rpt_family="MIR"	
repeat_region	3857..4202
rpt_family="MaLR"	
repeat_region	4664..4764
rpt_family="B4"	
repeat_region	5359..5661
rpt_family="MaLR"	
repeat_region	7140..7436
rpt_family="B4"	
repeat_region	9054..9193
rpt_family="Alu"	
repeat_region	9418..9597
rpt_family="L1"	
repeat_region	9616..10003
rpt_family="MaLR"	

```
repeat_region 10058..10228
/rpt_family="L1"
repeat_region 10292..10397
/rpt_family="Alu"
repeat_region 11077..11385
/rpt_family="ERV1"
repeat_region 13662..13738
/rpt_family="ID"
repeat_region 13755..13854
/rpt_family="B4"
repeat_region 14203..14415
/rpt_family="B2"
repeat_region 16232..16553
/rpt_family="ERVK"
repeat_region 16569..16677
/rpt_family="MIR"
repeat_region 17677..17770
/rpt_family="Alu"
repeat_region 19278..19380
/rpt_family="MIR"
repeat_region 21252..21618
/rpt_family="MaLR"
repeat_region 22089..22426
/rpt_family="MaLR"
repeat_region 23364..23729
/rpt_family="MaLR"
repeat_region 24357..24578
/rpt_family="L1"
repeat_region 24929..25066
/rpt_family="L1"
repeat_region 25417..25534
/rpt_family="B4"
repeat_region 25674..25952
/rpt_family="L1"
repeat_region 25956..26090
/rpt_family="Alu"
repeat_region 26497..26695
/rpt_family="L1"
repeat_region 26695..27696
/rpt_family="L1"
repeat_region 27730..28175
/rpt_family="L1"
repeat_region 29234..29488
/rpt_family="L1"
repeat_region 29481..29552
/rpt_family="L1"
repeat_region 37286..37401
/rpt_family="MaLR"
repeat_region 37993..38138
/rpt_family="L1"
repeat_region 38142..38504
/rpt_family="MaLR"
repeat_region 38509..39418
/rpt_family="L1"
repeat_region 40183..40290
/rpt_family="ERVK"
repeat_region 40867..41197
/rpt_family="MaLR"
repeat_region 41847..41941
/rpt_family="MIR"
repeat_region 42638..42723
/rpt_family="L1"
repeat_region 42724..42872
/rpt_family="Alu"
repeat_region 42873..42939
/rpt_family="L1"
repeat_region 42949..43245
/rpt_family="L1"
repeat_region 43292..43467
/rpt_family="L1"
repeat_region 43618..44010
/rpt_family="MaLR"
repeat_region 44588..44862
```

```
repeat_region /rpt_family="ERVK"
45594..45706
/rpt_family="L1"
repeat_region 45710..45951
/rpt_family="B4"
repeat_region 45949..45993
/rpt_family="B4"
repeat_region 46009..46309
/rpt_family="MaLR"
repeat_region 46506..46772
/rpt_family="L2"
repeat_region 48306..48489
/rpt_family="MER1_type"
repeat_region 48715..49119
/rpt_family="MaLR"
repeat_region 49429..49562
/rpt_family="Alu"
repeat_region 50440..50531

Query Match 84.0%; Score 16.8; DB 9; Length 197233;
Best Local Similarity 90.0%; Pred. No. 1.2e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GAACTGACCTTCACGCTT 20
Db 190199 GACACTGACCTTCACCTT 190218

RESULT 34
AC125460
LOCUS AC125460 205859 bp DNA linear ROD 27-NOV-2003
DEFINITION Mus musculus BAC clone RP24-448D10 from chromosome 15, complete
sequence.
ACCESSION AC125460
VERSION AC125460.4 GI:37361089
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE
AUTHORS Kruchowski, S. and Bielicki, L.
TITLE The sequence of Mus musculus BAC clone RP24-448D10
JOURNAL Unpublished (2001)
REFERENCE 2 (bases 1 to 205859)
AUTHORS Wilson, R.
TITLE Sequencing of Mus musculus
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 205859)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (26-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 205859)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (26-MAR-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 205859)
AUTHORS Wilson, R.K.
TITLE Direct Submission
JOURNAL Submitted (02-OCT-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 6 (bases 1 to 205859)
AUTHORS Wilson, R.
TITLE Direct Submission
JOURNAL Submitted (27-NOV-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Oct 2, 2003 this sequence version replaced gi:29244772.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
```



30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

**MAPPING INFORMATION:**  
Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

**SOURCE INFORMATION:**  
The RPCI-24 BAC Library has been constructed by Pieter de Jong and coworkers (<http://www.chori.org>) from male C57BL/6J mouse spleen and/or brain genomic DNA. The clone and detailed information can be obtained from Pieter de Jong and coworkers at <http://www.chori.org>

**NEIGHBORING SEQUENCE INFORMATION:**  
This sequence is the entire insert of the clone.  
Location/Qualifiers  
1. .211847  
    /organism="Mus musculus"  
    /mol\_type="genomic DNA"  
    /db\_xref="taxon:10090"  
    /chromosome="9"  
    /map="9"  
    /clone="RP24-499N24"  
    /clone\_lib="RPCI-24"  
    1. .665  
    /rpt\_family="ERVK"  
    636. .1436  
    /rpt\_family="ERVK"  
    835. .1027  
    /note="Sequence derived from PCR product of project DNA."  
    1411. .1516  
    /rpt\_family="ERVK"  
    1491. .1653  
    /rpt\_family="ERVK"  
    1654. .2067  
    /rpt\_family="ERVK"  
    2800. .3026  
    /rpt\_family="MaLR"  
    3754. .4119  
    /rpt\_family="MaLR"  
    7381. .7540  
    /rpt\_family="ERV1"  
    8934. .9216  
    /rpt\_family="MaLR"  
    10700. .10838  
    /rpt\_family="Alu"  
    11678. .11734  
    /rpt\_family="B4"  
    11997. .12108  
    /rpt\_family="B4"  
    12136. .12205  
    /rpt\_family="B4"  
    12628. .12677  
    /rpt\_family="ERV1"  
    13483. .13607  
    /rpt\_family="MER1\_type"  
    13684. .13750  
    /rpt\_family="ID"  
    15132. .15425  
    /rpt\_family="ERVL"  
    16996. .17259  
    /rpt\_family="MER2\_type"  
    18142. .18228  
    /rpt\_family="Alu"  
    20240. .20410  
    /rpt\_family="Alu"  
    21070. .21185  
    /rpt\_family="Alu"  
    21445. .21547

```

repeat_region /rpt_family="Alu"
22065..22247
repeat_region /rpt_family="L1"
22806..22974
repeat_region /rpt_family="MaLR"
23137..23522
repeat_region /rpt_family="MaLR"
26186..26196
repeat_region /rpt_family="B4"
26197..26335
repeat_region /rpt_family="Alu"
26336..26391
repeat_region /rpt_family="B4"
30311..30424
repeat_region /rpt_family="Alu"
30640..30840
repeat_region /rpt_family="B4"
30897..30941
repeat_region /rpt_family="Alu"
30914..30965
repeat_region /rpt_family="B4"
32853..33007
repeat_region /rpt_family="L1"
34531..34988
repeat_region /rpt_family="MaLR"
35001..35259
repeat_region /rpt_family="L1"
36957..37246
repeat_region /rpt_family="B4"
38434..38531
repeat_region /rpt_family="L1"
38660..39295
repeat_region /rpt_family="L1"
41533..41888
repeat_region /rpt_family="MaLR"
42325..42376
repeat_region /rpt_family="ERV1"
42402..42691
repeat_region /rpt_family="B4"
46821..46931
repeat_region /rpt_family="L1"
49087..49401
repeat_region /rpt_family="MaLR"
50170..50201
misc_feature /note="Sequence derived from PCR product of genomic DNA."
51146..51232
repeat_region /rpt_family="L1"
51237..51321
repeat_region /rpt_family="MaLR"
52846..53245
repeat_region /rpt_family="MaLR"
53822..53971
repeat_region /rpt_family="MaLR"
53999..55497
repeat_region /rpt_family="L1"
55527..55690
repeat_region /rpt_family="L1"
56747..56831
repeat_region /rpt_family="L1"
57497..57659
repeat_region /rpt_family="B2"
57677..57778
repeat_region /rpt_family="Alu"
59030..59220
repeat_region /rpt_family="B2"
59230..59584
repeat_region /rpt_family="L1"
71964..72054
repeat_region /rpt_family="Alu"
72944..73044
repeat_region /rpt_family="Alu"
73874..74005
repeat_region /rpt_family="MIR"

```

```

repeat_region 74109..74293
/rpt_family="B2"
repeat_region 79096..79263
/rpt_family="B4"

Query Match      84.0%; Score 16.8; DB 9; Length 211847;
Best Local Similarity 90.0%; Pred. No. 1.2e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 GAAACTGACCTTCACGCCTT 20
    |||||
Db 44887 GCATCTGACCTTCACGCCTT 44906
    |||||

RESULT 36
AC111633
LOCUS      Rattus norvegicus clone CH230-96P2, WORKING DRAFT SEQUENCE, 4
DEFINITION
AC111633
ACCESSION  AC111633.4 GI:25083873
VERSION    HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
KEYWORDS   Rattus norvegicus (Norway rat)
SOURCE     Rattus norvegicus
ORGANISM   Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Rattus;
MURNY,D.Marie., Metzker,M.Ilee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flaggs,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Guerra,W.,
Georgescu,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Loulisedge,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munkidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfankoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu.L.-L.,
Puazo,M., Quirroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,S., Song,X.-Z., Sorrelle,R., Sosa,J., Taylor,C.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,

```

Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

# TITLE JOURNAL REFERENCE AUTHORS TITLE JOURNAL

Unpublished  
2 (bases 1 to 214126)  
Worley, K.C.  
Direct Submission  
Submitted (19-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

# REFERENCE AUTHORS TITLE JOURNAL

3 (bases 1 to 214126)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

# COMMENT

On Nov 19, 2002 this sequence version replaced gi:23664598. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu

----- Project Information  
Center project name: GWMT  
Center clone name: CH230-96P2

----- Summary Statistics

Assembly program: Phrap; version 0.990329  
Consensus quality: 189317 bases at least Q40  
Consensus quality: 193026 bases at least Q30  
Consensus quality: 195308 bases at least Q20  
Estimated insert size: 196646; sum-of-contigs estimation  
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

-----  
\* 1 207709: contig of 207709 bp in length  
\* 207710 207809: gap of unknown length  
\* 207810 208862: contig of 1053 bp in length  
\* 208862 208962: gap of unknown length  
\* 208963 210348: contig of 1386 bp in length  
\* 210349 210448: gap of unknown length  
\* 210449 214126: contig of 3678 bp in length.

# FEATURES

Location/Qualifiers  
1. 214126  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-96P2"  
/note="clone boundary  
587. 1070  
clone\_end:T7"

misc\_feature  
gap  
gap  
gap  
ORIGIN

Query Match 84.0%; Score 16.8; DB 14; Length 214126;  
Best Local Similarity 90.0%; Pred. No. 1.2e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCAGCCTT 20  
|||||  
Db 195638 GAAACTGACCTTCAGCCTT 195657  
|||||

RESULT 37  
AC128057

LOCUS  
DEFINITION  
AC128057.3 GI:25074509  
HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
KEYWORDS  
SOURCE  
Rattus norvegicus  
Rattus norvegicus  
ORGANISM

REFERENCE  
AUTHORS

1 (bases 1 to 216428)  
Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J., Allen, C., Allent, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Chen, Y., Chen, Y., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Y., Chu, J., Cleaveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Gear, K., Gili, R., Grady, M., Guerra, W., Guevara, M., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Haviak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Loulseghe, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhiney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokelemeh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Pu, L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,

Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Suton,A., Svatek,A., Tabot,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleciyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,P., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

## TITLE

Direct Submission

## REFERENCE

Unpublished

## AUTHORS

Worley,K.C.

## TITLE

Direct Submission

## JOURNAL

Submitted (19-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

## REFERENCE

3 (bases 1 to 216428)

## AUTHORS

Rat Genome Sequencing Consortium.

## JOURNAL

Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 19, 2002 this sequence version replaced gi:23665158. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information

Center project name: KAT

Center clone name: CH230-17111

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 207166 bases at least Q40

Consensus quality: 208862 bases at least Q30

Consensus quality: 209986 bases at least Q20

Estimated insert size: 210133; sum-of-contigs estimation

Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 214801: contig of 214801 bp in length

\* 214802 214901: gap of unknown length

\* 214902 216428: contig of 1527 bp in length.

## FEATURES

source

Location/Qualifiers

1. .216428

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clone="CH230-17111"

misc\_feature

1. .1048

/note="wgs end\_extension

clone end:T7"

complement(3743..4501)

/note="clone\_boundary

clone\_end:T7

site:

end sequence:BH278612"

211695..212414

/note="clone\_boundary

clone\_end:Sp6

site:

end sequence:BH278642"

213638..214801

/note="wgs end\_extension

clone end:Sp6"

214802..214901

/estimated\_length=unknown

## ORIGIN

Query Match 84.0%; Score 16.8; DB 14; Length 216428;  
 Best Local Similarity 90.0%; Pred. No. 1.2e+03;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GAAACTGACCTTCAGCCTT 20  
 ||||| ||||| ||||| |||||  
 Db 134262 GAAACTGACTTTCATGCTT 134281

## RESULT 38

AC162547/c

LOCUS

AC162547

ACCESSION

AC162547.2

VERSION

HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.

KEYWORDS

Bos taurus (cow)

SOURCE

Bos taurus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;

Pecora; Bovidae; Bovinae; Bos.

1 (bases 1 to 229549)

REFERENCE

AUTHORS

Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Delgado,M.B., Davis,C., Davy-Carroil,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Evans,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hennings,B., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuewa,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Muniadasa,M., Murphy,M., Nair,L.,





```

/estimated length=50
210978..211094
/estimated length=117
224662..224761
/estimated length=unknown
225978..226077
/estimated length=unknown
227481..227580
/estimated length=unknown

```

## ORIGIN

```

Query Match      84.0%; Score 16.8; DB 14; Length 229549;
Best Local Similarity 90.0%; Pred. No. 1.2e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 GAAACTGACCTTCAGCCCTT 20
|| ||||| ||||| ||||| ||
Db 58753 GACACTGACCTTCAGCCCT 58734

```

```

RESULT 39
AC105577
LOCUS      239066 bp      DNA      linear      HTG 09-NOV-2002
DEFINITION Rattus norvegicus clone CH230-233D22, WORKING DRAFT SEQUENCE, 3
unordered pieces.
ACCESSION AC105577
VERSION   AC105577.4 GI:24818789
KEYWORDS  HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE    Rattus norvegicus (Norway rat)
ORGANISM  Rattus norvegicus

```

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurgnathi; Muridae; Murinae; Rattus.

```

```

1 (bases 1 to 239066)
Muzny,D.Marie., Metzker,M.Lee., Abranzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,

```

```

Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,P.,
Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,X., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunarathne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,X., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorenshewa,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapa,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwakkilemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfankoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu.L.-L.,
Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savary,G., Scherer,S., Scott,G., Shatman,S., Shen,H.,
Shetty,J., Shvartebeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,

```

```

Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wlezyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.

```

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

```

/misc_feature
AC100736/c
  /clone="CH230-233D22"
  complement(31483..32335)
  /note="clone_boundary
  clone_end:T7
  site:ECORI
  end_sequence:BZ093528"
  33399..33498
  /estimated_length=unknown
  complement(196731..197572)
  /note="clone_boundary
  clone_end:Sp6
  site:ECORI
  end_sequence:BZ093530"
  235146..235245
  /estimated_length=unknown

ORIGIN
Query Match      84.0%; Score 16.8; DB 14; Length 239066;
Best Local Similarity 90.0%; Pred. No. 1.2e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACGCCTT 20
   |||||
Db 109473 GACACTGACCTTCATGCCTT 109492

RESULT 40
AC100736/c
LOCUS
DEFINITION Mus musculus chromosome 15, clone RP24-33606, complete sequence.
ACCESSION AC100736
VERSION AC100736.5 GI:33354406
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 247275)
Birren,B., Nusbaum,C. and Lander,E.
Mus musculus chromosome 15, clone RP24-33606
Unpublished
2 (bases 1 to 247275)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepel,Y., Colangelo,M., Collins,S., Collumore,A., Cook,A.,
Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosett,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 247275)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collumore,A., Cook,A., Cooke,P., Corum,B., DeArelano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., Maclean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ranasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (30-JUL-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 247275)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collumore,A., Cook,A., Cooke,P., Corum,B., DeArelano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., Maclean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ranasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (30-JUL-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
5 (bases 1 to 247275)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collumore,A., Cook,A., Cooke,P., Corum,B., DeArelano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., Maclean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ranasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (31-JUL-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
6 (bases 1 to 247275)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collumore,A., Cook,A., Cooke,P., Corum,B., DeArelano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., Maclean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ranasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (30-SEP-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research

```

Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequencing\\_submission@genome.wi.mit.edu](mailto:sequencing_submission@genome.wi.mit.edu)  
----- Project Information  
Center project name: L17452  
Center clone name: 336\_O\_6  
-----

FEATURES

source  
1. .247275  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="15"  
/map="15"  
/clone="RP24-33608"  
/clone\_lib="RPCI-24 Male Mouse BAC"  
1. .27236  
/note="wgs end extension  
clone end:Sp6"  
complement(380..470)  
/rpt\_family="B4A"  
731..756  
/rpt\_family="AT rich"  
complement(1231..1368)  
/rpt\_family="Lx7"  
1379..1432  
/rpt\_family="AT rich"  
complement(1445..2616)  
/rpt\_family="Lx7"  
2332..2366  
/note="single clone coverage"  
complement(2612..4009)  
/rpt\_family="Lx7"  
4420..4448  
/rpt\_family=" (TG)n"  
4628..4955  
/rpt\_family="MTD"  
5106..5137  
/rpt\_family=" (TTTA)n"  
5160..5179  
/rpt\_family=" (TG)n"  
complement(5181..5338)  
/rpt\_family="RSINE1"  
5373..5399  
/rpt\_family=" (TG)n"  
complement(5406..5577)  
/rpt\_family="LIM4"  
5918..5932  
/note="<30 qual SNGL region"  
5944..5987  
/note="<30 qual SNGL region"  
6219..6300  
/rpt\_family="MTD"  
6327..6402  
/rpt\_family=" (CA)n"  
6404..6456  
/rpt\_family=" (TCTA)n"  
6459..6526  
/rpt\_family=" (TC)n"  
6529..6568  
/rpt\_family=" (TG)n"  
complement(6569..6983)  
/rpt\_family="MT2A"  
6992..7252  
/rpt\_family="MTE"  
7312..7335  
/rpt\_family="AT rich"  
7439..7485  
/rpt\_family="AT rich"  
7675..7699  
/rpt\_family=" (TG)n"  
7788..8360  
/rpt\_family="RMER1B"

repeat\_region 8361..8393  
/rpt\_family=" (TTTTA)n"  
repeat\_region 8394..8506  
/rpt\_family="RMER1B"  
repeat\_region complement(8717..8818)  
/rpt\_family="MTE"  
repeat\_region 8819..8888  
/rpt\_family=" (GA)n"  
repeat\_region complement(8889..8924)

Query Match 84.0%; Score 16.8; DB 9; Length 247275;  
Best Local Similarity 90.0%; Pred. No. 1.2e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GAAACTGACCTTCAGGCCTT 20  
|||||  
Db 210513 GAAACTGAACCTTCAGGCCTT 210494

Search completed: November 20, 2005, 18:33:25  
Job time : 739.099 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2005, 15:03:55 ; Search time 324.938 Seconds  
(without alignments)  
410.213 Million cell updates/sec

Title: US-10-627-757-21

Perfect score: 20  
Sequence: 1 gaaactgaccttcagccctt 20

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_21:\*  
1: Geneseqn1980s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002as:\*  
7: Geneseqn2002bs:\*  
8: Geneseqn2003as:\*  
9: Geneseqn2003bs:\*  
10: Geneseqn2003cs:\*  
11: Geneseqn2003ds:\*  
12: Geneseqn2004as:\*  
13: Geneseqn2004bs:\*  
14: Geneseqn2005s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	20	12	ADL14969
2	20	100.0	20	14	ADW86591
3	20	100.0	1074	12	ADL14953
4	20	100.0	1074	14	ADM86575
5	20	100.0	46951	10	ADE13891
c 6	16.8	84.0	891	13	ADS55422
c 7	16.8	84.0	891	13	ADU85916
8	16.8	84.0	2736	8	AB212430
9	16.8	84.0	2736	8	AB242033
10	16.8	84.0	3105	4	AAI66065
11	16.8	84.0	6376	4	AAI66066
c 12	16.4	82.0	3213	8	ACA54251
c 13	16	80.0	596	13	ADW85916
14	16	80.0	2751	13	ADU07712
15	15.8	79.0	241	13	ACF85011
16	15.8	79.0	299	8	AB256651
17	15.8	79.0	381	8	AB256847
18	15.8	79.0	382	8	AB256844
19	15.8	79.0	440	8	ABQ83840

20	15.8	79.0	440	8	ABV77524
21	15.8	79.0	497	8	ABZ56877
22	15.8	79.0	510	7	AD31337
23	15.8	79.0	510	7	ADY36725
24	15.8	79.0	522	8	ABZ56884
25	15.8	79.0	641	3	AAF08072
26	15.8	79.0	641	13	ADU52113
27	15.8	79.0	641	14	ADZ90116
28	15.8	79.0	1173	10	ACF68857
29	15.8	79.0	1362	3	AAF13166
30	15.8	79.0	1362	13	ADU57207
31	15.8	79.0	1362	14	ADZ95210
c 32	15.8	79.0	1476	8	ACA00487
c 33	15.8	79.0	1491	5	AAH66345
c 34	15.8	79.0	1621	10	ADD13197
c 35	15.8	79.0	2652	11	ACL29777
c 36	15.8	79.0	5368	6	ABN80098
c 37	15.8	79.0	5882	6	ABL32545
38	15.8	79.0	38538	13	ABD33357
c 39	15.8	79.0	110000	10	ACF67367
c 40	15.8	79.0	209613	14	ADY25743
c 41	15.8	79.0	249878	10	ACF65381
c 42	15.8	79.0	349980	5	AAH68528
c 43	15.4	77.0	460	6	ABK76807
c 44	15.4	77.0	624	13	ADQ55239
45	15.4	77.0	707	3	AAC74780

ALIGNMENTS

RESULT 1

ADL14969  
ID ADL14969 standard; DNA; 20 BP.  
XX  
AC ADL14969;  
XX  
DT 06-MAY-2004 (first entry)  
XX  
DE Human glaucoma-related optineurin (OPTN) exon 7 PCR primer SF7.  
XX  
KW Human; glaucoma; optineurin; OPTN; diagnosis; PCR; primer; ss.  
XX  
OS Homo sapiens.  
XX  
PN EPI388590-A2.  
XX  
PD 11-FEB-2004.  
XX  
PF 29-JUL-2003; 2003EP-00447201.  
XX  
PR 02-AUG-2002; 2002JP-00226612.  
XX  
PA (SYSM-) SYSMEX CORP.

XX Kouchi Y, Masago A, Takahata T;  
XX WPI; 2004-146134/15.  
XX  
XX Gene assay for predicting future onset of glaucoma, particularly primary  
XX open angle glaucoma or normal ocular tension glaucoma, comprises  
XX detecting a mutation of at least one base of the optineurin gene.  
XX  
XX Claim 9; SEQ ID NO 21; 31pp; English.

XX The present sequence is that of PCR primer SF7 for exon 7 ADL14953 of the  
XX glaucoma-associated gene, OPTN (optineurin) ADL14949. The invention  
XX relates to a gene assay method for predicting future onset of primary  
XX open angle glaucoma and/or normal ocular tension glaucoma. This involves  
XX detecting a mutation in the OPTN gene coding sequence, specifically a  
XX substitution of G for A at position 619 and/or a substitution of A for G  
XX at position 998 of the OPTN coding sequence. The mutation(s) is detected  
XX using a nucleic acid amplification method using primers specific for the

```
CC different exons of the coding sequence, including primers SF7 and SR7
CC ADL14970 for exon 7.
XX
SQ Sequence 20 BP; 5 A; 7 C; 3 G; 5 T; 0 U; 0 Other;

Query Match          100.0%; Score 20; DB 12; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACGCCTT 20
Db 1 GAAACTGACCTTCACGCCTT 20

RESULT 2
ADW86591
ID ADW86591 standard; DNA; 20 BP.
XX
AC ADW86591;
XX
DT 21-APR-2005 (first entry)
XX
DE PCR primer used to amplify human optineurin (OPTN) gene exon 7 Seq 21.
XX
KW glaucoma; optineurin; ophthalmological; ss; ocular disease;
KW DNA amplification; genetic marker; PCR; primer.
XX
OS Homo sapiens.
XX
PN JP2005034112-A.
XX
PD 10-FEB-2005.
XX
PF 29-JUL-2003; 2003JP-00281897.
XX
PR 02-AUG-2002; 2002JP-00226612.
XX
PR 30-JUN-2003; 2003JP-00188070.
XX
PA (TOAI-) TOA IYO DENSHI KK.
XX
WPI; 2005-156038/17.
XX
Estimating risk of onset of glaucoma, involves analyzing mutation in any
PT one portion of Optic new phosphorous gene, and utilizing analyzed
PT mutation as index for estimating risk of onset of glaucoma.
XX
Example 1; SEQ ID NO 21; 13pp; Japanese.
XX
This invention relates to a novel method for estimating the risk of onset
CC of glaucoma. Specifically, it refers to a method that involves analyzing
CC a mutation in any one portion of an optineurin (OPTN) gene and utilizing
CC the analyzed mutation as an index for estimating the risk of onset of
CC glaucoma. The present invention describes oligonucleotides to detect
CC mutations that hybridize with one or more portions of the OPTN glaucoma
CC related gene. Accordingly, it provides oligos that detect A619G and G898A
CC mutations in the human OPTN gene of patients at risk of developing
CC glaucoma, in particular primary open-angle glaucoma and/or normal tension
CC glaucoma. These mutations will not be present in a sample obtained from a
CC healthy person who is not at risk of onset of glaucoma. This
CC oligonucleotide sequence is a PCR primer used to amplify a human OPTN
CC exon of the invention.
XX
SQ Sequence 20 BP; 5 A; 7 C; 3 G; 5 T; 0 U; 0 Other;

Query Match          100.0%; Score 20; DB 14; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACGCCTT 20
Db 1 GAAACTGACCTTCACGCCTT 20
```

```
RESULT 3
ADL14953
ID ADL14953 standard; DNA; 1074 BP.
XX
AC ADL14953;
XX
DT 06-MAY-2004 (first entry)
XX
DE Human glaucoma-related optineurin (OPTN) exon 7.
XX
KW Human; glaucoma; optineurin; OPTN; diagnosis; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT primer_bind complement (261..280)
FT /*tag= a
FT /*note= "Primer SF7"
FT exon 501..574
FT /*tag= b
FT /*number= 7
FT primer_bind 765..784
FT /*tag= c
FT /*note= "Primer SR7"
XX
PN EP1388590-A2.
XX
PD 11-FEB-2004.
XX
PF 29-JUL-2003; 2003EP-00447201.
XX
PR 02-AUG-2002; 2002JP-00226612.
XX
PA (SYSM-) SYSMEX CORP.
XX
PI Kouchi Y, Masago A, Takahata T;
XX
WPI; 2004-146134/15.
XX
Gene assay for predicting future onset of glaucoma, particularly primary
PT open angle glaucoma or normal ocular tension glaucoma, comprises
PT detecting a mutation of at least one base of the optineurin gene.
XX
Example 1; SEQ ID NO 5; 31pp; English.
XX
The present sequence comprises exon 7 of the glaucoma-associated gene,
CC OPTN (optineurin) ADL14949. The invention relates to a gene assay method
CC for predicting future onset of primary open angle glaucoma and/or normal
CC ocular tension glaucoma. This involves detecting a mutation in the OPTN
CC gene coding sequence, specifically a substitution of G for A at position
CC 619 and/or a substitution of A for G at position 898 of the OPTN coding
CC sequence. The mutation(s) is detected using a nucleic acid amplification
CC method using primers specific for the different exons of the coding
CC sequence, including primers SF7 ADL14969 and SR7 ADL14970 for exon 7.
XX
SQ Sequence 1074 BP; 279 A; 222 C; 249 G; 324 T; 0 U; 0 Other;

Query Match          100.0%; Score 20; DB 12; Length 1074;
Best Local Similarity 100.0%; Pred. No. 3.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACGCCTT 20
Db 261 GAAACTGACCTTCACGCCTT 280

RESULT 4
ADW86575
ID ADW86575 standard; DNA; 1074 BP.
XX
AC ADW86575;
XX
DT 21-APR-2005 (first entry)
```

```

XX DE Human optineurin (OPTN) gene exon 7 Seq 5.
XX KW glaucoma; optineurin; ophthalmological; ds; ocular disease;
KW DNA amplification; genetic marker.
XX OS Homo sapiens.
XX PN JP2005034112-A.
XX PD 10-FEB-2005.
XX PF 29-JUL-2003; 2003JP-00281897.
XX PR 02-AUG-2002; 2002JP-00226612.
XX PR 30-JUN-2003; 2003JP-00188070.
XX PA (TOAI-) TOA IYO DENSHI KK.
XX DR WPI; 2005-156038/17.
XX PF Estimating risk of onset of glaucoma, involves analyzing mutation in any
XX PT one portion of Optic new phosphorous gene, and utilizing analyzed
XX PT mutation as index for estimating risk of onset of glaucoma.
XX PS Example 1; SEQ ID NO 5; 13pp; Japanese.
XX CC This invention relates to a novel method for estimating the risk of onset
XX CC of glaucoma. Specifically, it refers to a method that involves analyzing
XX CC a mutation in any one portion of an optineurin (OPTN) gene and utilizing
XX CC the analyzed mutation as an index for estimating the risk of onset of
XX CC glaucoma. The present invention describes oligonucleotides to detect
XX CC mutations that hybridize with one or more portions of the OPTN glaucoma
XX CC related gene. Accordingly, it provides oligos that detect A619G and G898A
XX CC mutations in the human OPTN gene of patients at risk of developing
XX CC glaucoma, in particular primary open-angle glaucoma and/or normal tension
XX CC glaucoma. These mutations will not be present in a sample obtained from a
XX CC healthy person who is not at risk of onset of glaucoma. This
XX CC polynucleotide sequence is a human OPTN exon of the invention.
XX SQ Sequence 1074 BP; 279 A; 222 C; 249 G; 324 T; 0 U; 0 Other;

Query Match 100.0%; Score 20; DB 14; Length 1074;
Best Local Similarity 100.0%; Pred. No. 3.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACGCGCTT 20
Db 261 GAAACTGACCTTCACGCGCTT 280

RESULT 5
ADE13891
ID ADE13891 standard; DNA; 46951 BP.
AC ADE13891;
XX 29-JAN-2004 (first entry)
XX DE Human optineurin gene.
XX KW Human; optineurin; ds; gene; ophthalmological;
KW single nucleotide polymorphism; SNP; glaucoma;
KW progressive ocular hypertensive disorder; glaucoma related disorder.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX variation replace(391,G)
XX FT /tag= a
XX FT /standard_name= "Single nucleotide polymorphism"
XX FT replace(691,G)
XX FT /tag= b

```

```

FT variation /standard_name= "Single nucleotide polymorphism"
FT replace(709,G)
FT /tag= c
FT variation /standard_name= "Single nucleotide polymorphism"
FT replace(887,A)
FT /tag= d
FT variation /standard_name= "Single nucleotide polymorphism"
FT replace(894,T)
FT /tag= e
FT variation /standard_name= "Single nucleotide polymorphism"
FT replace(987,C)
FT /tag= f
FT variation /standard_name= "Single nucleotide polymorphism"
FT replace(1112,C)
FT /tag= g
FT variation /standard_name= "Single nucleotide polymorphism"
FT replace(1505,CC)
FT /tag= h
FT variation /standard_name= "Single nucleotide polymorphism"
FT replace(1606,A)
FT /tag= i
FT variation /standard_name= "Single nucleotide polymorphism"
FT replace(2405,T)
FT /tag= j
FT variation /standard_name= "Single nucleotide polymorphism"
FT replace(2606,G)
FT /tag= k
FT variation /standard_name= "Single nucleotide polymorphism"
FT replace(3313,A)
FT /tag= l
FT variation /standard_name= "Single nucleotide polymorphism"
FT replace(3555,TT)
FT /tag= m
FT variation /standard_name= "Single nucleotide polymorphism"
FT replace(3625,G)
FT /tag= n
FT variation /standard_name= "Single nucleotide polymorphism"
FT replace(3629,C)
FT /tag= o
FT variation /standard_name= "Single nucleotide polymorphism"
FT replace(3882,TT)
FT /tag= p
FT variation /standard_name= "Single nucleotide polymorphism"
FT replace(3988,T)
FT /tag= q
FT variation /standard_name= "Single nucleotide polymorphism"
FT replace(4452,A)
FT /tag= r
FT /standard_name= "Single nucleotide polymorphism"
XX US2003190617-A1.
XX PN 09-OCT-2003.
XX PD 06-MAR-2002; 2002US-00091281.
XX PF 06-MAR-2002; 2002US-00091281.
XX PR 06-MAR-2002; 2002US-00091281.
XX XX (SIEE/) SI E.
XX PA (RAYM/) RAYMOND V.
XX PA (MORI/) MORISSETTE J.
XX PI Raymond V, Morissette J, Si E;
XX WPI; 2003-864168/80.
XX DR New nucleic acid sequences of the optineurin gene are useful to detect
XX PT polymorphisms particularly single nucleotide polymorphisms in the
XX PT optineurin promoter to diagnose, prognosis and treat glaucoma and related
XX PS disorders.
XX PS Disclosure; SEQ ID NO 2; 159pp; English.
XX PS

```



CC The invention relates to an isolated nucleic acid (N1) comprising at  
 CC least 20 but not more than 1500 consecutive nucleotides of the optineurin  
 CC promoter appearing as ADE13890. Also included are the optineurin promoter  
 CC operably linked to a heterologous nucleic acid, a nucleic acid capable of  
 CC detecting a single nucleotide polymorphism (SNP) in the optineurin  
 CC promoter, a host cell comprising the promoter operably linked to a  
 CC heterologous sequence, diagnosing or prognosing glaucoma in a sample  
 CC obtained from a cell or bodily fluid (comprising detecting a polymorphism  
 CC in a promoter region of the optineurin gene, associated with a glaucoma  
 CC phenotype), detecting a SNP sequence variation in a sample containing  
 CC DNA, detecting the presence of an optineurin promoter sequence variation  
 CC in a sample containing DNA, determining the presence or increased  
 CC susceptibility to glaucoma or to a progressive ocular hypertensive  
 CC disorder resulting in loss of visual field in a patient for the severity  
 CC or progression of glaucoma in a patient, comprising providing  
 CC amplification reaction primers that direct amplification of a selected  
 CC nucleic acid region containing the variation within the optineurin  
 CC promoter and amplifying the DNA) and detecting a polymorphism (comprising  
 CC obtaining a sample containing human genomic DNA, providing a nucleic acid  
 CC capable of detecting a SNP located within an optineurin promoter, and  
 CC detecting the polymorphism). The invention is used to diagnose and  
 CC prognose glaucoma and also to treat glaucoma related disorders. The  
 CC present sequence is the optineurin gene.

SQ Sequence 46951 BP; 12703 A; 10108 C; 10051 G; 14089 T; 0 U; 0 Other;

Query Match 100.0%; Score 20; DB 10; Length 46951;  
 Best Local Similarity 100.0%; Pred. No. 6.5;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACGCCTT 20

DB 20871 GAAACTGACCTTCACGCCTT 20890

#### RESULT 6

ID ADS55422/c  
 AC ADS55422;

02-DEC-2004 (first entry)

Bacterial polynucleotide #7409.

Recombinant DNA construct; transformed plant; improved plant property;  
 cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;  
 pathogen tolerance; pest tolerance; plant disease resistance;  
 cell cycle pathway modification; plant growth regulator;  
 homologous recombination; seed oil yield; protein yield; carbohydrate;  
 nitrogen; phosphorus; photosynthesis; lignin; galactomannan;  
 bacterial polynucleotide; gene; ss.

OS Bacteria.

US2003233675-A1.

18-DEC-2003.

20-FEB-2003; 2003US-00369493.

21-FEB-2002; 2002US-0360039P.

(CAOY/) CAO Y.  
 (HINK/) HINKLE G J.  
 (SLAT/) SLATER S C.  
 (CHEN/) CHEN X.  
 (GOLD/) GOLDMAN B S.

CAO Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;  
 WPI; 2004-061375/06.

PT New recombinant DNA construct comprising a promoter positioned to provide  
 PT for expression of a polynucleotide encoding a polypeptide from a  
 PT microbial source, useful for producing plants with improved properties.

PS Claim 1; SEQ ID NO 31096; 122pp; English.

CC The invention relates to a recombinant DNA construct comprising a  
 CC promoter functional in a plant cell, where the promoter is positioned to  
 CC provide for expression of a polynucleotide encoding a polypeptide from a  
 CC microbial source. The invention also relates to a transformed plant  
 CC comprising the recombinant DNA construct and a method of producing a  
 CC transformed plant having an improved property. The plant is a crop plant  
 CC such as maize or soybean. The method of producing a transformed plant  
 CC having an improved property comprises transforming a plant with the  
 CC recombinant DNA construct and growing the transformed plant, where the  
 CC polynucleotide or polypeptide is useful for improving plant properties.  
 CC The recombinant DNA construct is useful for producing plants with  
 CC improved plant properties, e.g. improved cold, heat or drought tolerance,  
 CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,  
 CC increased resistance to plant disease, better growth rate by modification  
 CC of the cell cycle pathway with plant growth regulators, increased rate of  
 CC homologous recombination, modified seed oil or protein yield and/or  
 CC content, improved yield by modification of carbohydrate, nitrogen or  
 CC phosphorus use and/or uptake, by modification of photosynthesis or by  
 CC providing improved plant growth and development under at least one stress  
 CC condition, improved lignin production or improved galactomannan  
 CC production. This sequence represents a bacterial polynucleotide used in  
 CC the scope of the invention. Note: The sequence data for this patent did  
 CC not form part of the printed specification but was obtained in electronic  
 CC format from USPTO at seqdata.uspto.gov/sequence.html.

SQ Sequence 891 BP; 132 A; 321 C; 297 G; 141 T; 0 U; 0 Other;

Query Match 84.0%; Score 16.8; DB 13; Length 891;  
 Best Local Similarity 90.0%; Pred. No. 1.7e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACGCCTT 20

DB 465 GAAACTGATTTTCACGCCTT 446

#### RESULT 7

ADS49908/c  
 ID ADS49908 standard; cDNA; 891 BP.

AC ADS49908;

02-DEC-2004 (first entry)

Bacterial polynucleotide #4651.

Recombinant DNA construct; transformed plant; improved plant property;  
 cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;  
 pathogen tolerance; pest tolerance; plant disease resistance;  
 cell cycle pathway modification; plant growth regulator;  
 homologous recombination; seed oil yield; protein yield; carbohydrate;  
 nitrogen; phosphorus; photosynthesis; lignin; galactomannan;  
 bacterial polynucleotide; gene; ss.

OS Bacteria.

US2003233675-A1.

18-DEC-2003.

20-FEB-2003; 2003US-00369493.

21-FEB-2002; 2002US-0360039P.

(CAOY/) CAO Y.  
 (HINK/) HINKLE G J.  
 (SLAT/) SLATER S C.

PA (CHEN/) CHEN X.  
 XX (GOLD/) GOLDMAN B S.  
 XX  
 PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;  
 XX WPI; 2004-061375/06.  
 DR  
 XX  
 PT New recombinant DNA construct comprising a promoter positioned to provide  
 PT for expression of a polynucleotide encoding a polypeptide from a  
 PT microbial source, useful for producing plants with improved properties.  
 XX  
 XX Claim 1; SEQ ID NO 28338; 122pp; English.  
 XX  
 CC The invention relates to a recombinant DNA construct comprising a  
 CC promoter functional in a plant cell, where the promoter is positioned to  
 CC provide for expression of a polynucleotide encoding a polypeptide from a  
 CC microbial source. The invention also relates to a transformed plant  
 CC comprising the recombinant DNA construct and a method of producing a  
 CC transformed plant having an improved property. The plant is a crop plant  
 CC such as maize or soybean. The method of producing a transformed plant  
 CC having an improved property comprises transforming a plant with the  
 CC recombinant DNA construct and growing the transformed plant, where the  
 CC polynucleotide or polypeptide is useful for improving plant properties.  
 CC The recombinant DNA construct is useful for producing plants with  
 CC improved plant properties, e.g. improved cold, heat or drought tolerance,  
 CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,  
 CC increased resistance to plant disease, better growth rate by modification  
 CC of the cell cycle pathway with plant growth regulators, increased rate of  
 CC homologous recombination, modified seed oil or protein yield and/or  
 CC content, improved yield by modification of carbohydrate, nitrogen or  
 CC phosphorus use and/or uptake, by modification of photosynthesis or by  
 CC providing improved plant growth and development under at least one stress  
 CC condition, improved lignin production or improved galactomannan  
 CC production. This sequence represents a bacterial polynucleotide used in  
 CC the scope of the invention. Note: The sequence data for this patent did  
 CC not form part of the printed specification but was obtained in electronic  
 CC format from USPTO at seqdata.uspto.gov/sequence.html.  
 XX  
 SQ Sequence 891 BP; 132 A; 321 C; 297 G; 141 T; 0 U; 0 Other;  
 Query Match 84.0%; Score 16.8; DB 13; Length 891;  
 Best Local Similarity 90.0%; Pred. No. 1.7e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 GAAACTGACCTTCACGCGCTT 20  
 |||||  
 Db 465 GAAACTGATTTTCACGCGCTT 446  
 RESULT 8  
 ABZ12430  
 ID ABZ12430 standard; DNA; 2736 BP.  
 AC  
 XX ABZ12430;  
 XX  
 DT 21-JAN-2003 (first entry)  
 XX  
 DE Arabidopsis thaliana stress regulated gene SEQ ID NO 235.  
 XX  
 KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 PN WO200216655-A2.  
 XX  
 PD 28-FEB-2002.  
 XX  
 XX 24-AUG-2001; 2001WO-US026685.  
 PF  
 XX  
 XX 24-AUG-2000; 2000US-0227866P.  
 PR  
 XX 26-JAN-2001; 2001US-0264647P.  
 PR  
 XX 22-JUN-2001; 2001US-0300111P.  
 XX

PA (SCRI ) SCRIPPS RES INST.  
 PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
 XX  
 PI Harper JF, Krops J, Wang X, Zhu T;  
 XX WPI; 2002-304127/34.  
 DR  
 XX  
 PT Identifying a stress condition to which a plant cell has been exposed and  
 PT producing plants with increased tolerance to these abiotic stresses.  
 XX  
 XX Claim 144; SEQ ID NO 235; 577pp + Sequence Listing; English.  
 XX  
 CC The invention relates to identifying a stress condition to which a plant  
 CC cell has been exposed, comprising: (a) contacting nucleic acid  
 CC representative of expressed polynucleotides in the plant cell with an  
 CC array or probes representative of the plant cell genome; and (b)  
 CC detecting a profile of expressed polynucleotides in the plant cell  
 CC characteristic of a stress response. The method is useful in the  
 CC production of transgenic plants, cells and seeds and in producing plants  
 CC with increased tolerance to abiotic stress. The present sequence is that  
 CC of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used  
 CC in methods of the invention. Note: The sequence data for this patent is  
 CC not represented in the printed specification but is based on sequence  
 CC information supplied to Derwent by the European Patent Office  
 XX  
 SQ Sequence 2736 BP; 784 A; 481 C; 785 G; 686 T; 0 U; 0 Other;  
 Query Match 84.0%; Score 16.8; DB 6; Length 2736;  
 Best Local Similarity 90.0%; Pred. No. 2e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 GAAACTGACCTTCACGCGCTT 20  
 |||||  
 Db 1270 GAAATTGAATTCACGCGCTT 1289  
 RESULT 9  
 ABZ42033  
 ID ABZ42033 standard; cDNA; 2736 BP.  
 XX  
 AC ABZ42033;  
 XX  
 DT 27-FEB-2003 (first entry)  
 XX  
 DE Arabidopsis thaliana gene #17 modulated by PTGS.  
 XX  
 KW Posttranscriptional gene silencing; PTGS; plant; transfection; gene;  
 KW ss.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..2736  
 FT /\*tag= a  
 XX  
 XX WO200281695-A2.  
 XX  
 PD 17-OCT-2002.  
 XX  
 PF 05-APR-2002; 2002WO-EP003806.  
 XX  
 PR 06-APR-2001; 2001US-0282049P.  
 XX  
 PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
 PA (FRIE-) FRIEDRICH MIESCHER INST.  
 XX  
 XX Zhu T, Glazov EA, Meins F, Wang X, Chang H;  
 XX WPI; 2003-103337/09.  
 DR P-PSDB; ABP81189.  
 XX  
 XX Novel polynucleic acid segment useful for modulating gene expression  
 PT within a cell by posttranscriptional gene silencing, and for augmenting a

```

PT plant cell genome.
XX Claim 18; Page 185-186; 439pp; English.
XX The invention relates to a novel isolated polynucleic acid segment
CC modulated within a cell by posttranscriptional gene silencing (PTGS). The
CC invention specifically relates to a method to identify an expression
CC product that is modulated by PTGS. The polynucleotide is useful for
CC modulating the gene expression within a cell by PTGS, by introducing the
CC polynucleic acid into a cell and expressing the nucleic acid segment in
CC the cell to form a product. The polynucleic acid segment is also useful
CC for augmenting a cell genome, and for augmenting a plant genome, by
CC contacting a plant cell with the segment to produce a transformed plant
CC cell, and growing the transformed plant cell to produce a differentiated
CC transformed plant. The sequences shown in ABZ42017 - ABZ42142 represent
CC segments of A. thaliana cDNA modulated by PTGS
XX Sequence 2736 BP; 784 A; 481 C; 785 G; 686 T; 0 U; 0 Other;
SQ Query Match 84.0%; Score 16.8; DB 8; Length 2736;
Best Local Similarity 90.0%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GAAACTGACCTTCACGCCTT 20
Db 1270 GAAATTGAACCTTCACGCCTT 1289

RESULT 10
AAI66065
ID AAI66065 standard; DNA; 3105 BP.
XX AAI66065;
XX 14-JAN-2002 (first entry)
XX Arabidopsis thaliana heat shock protein 101 gene (GenBank: U13949).
XX Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
XX canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
XX Arabidopsis thaliana; ds.
XX Arabidopsis thaliana.
XX WO200170929-A2.
XX 27-SEP-2001.
XX 20-MAR-2001; 2001WO-US008836.
XX 20-MAR-2000; 2000US-0190769P.
XX 18-APR-2000; 2000US-0198116P.
XX (ARCH-) ARCH DEV CORP.
XX Lindquist S, Queitsch C, Vierling E;
XX WPI; 2001-639123/73.
XX P-PSDB; AAM51659.
XX Transgenic plants with improved heat stress tolerance, useful for
XX producing animal feed, oil and synthetic products.
XX Claim 4; Page; 91pp; English.
XX The invention relates to a transgenic plant, comprising a genetic
XX construct comprising a promoter operatively linked to a nucleic acid
XX sequence (AAI66057-AAI66084) encoding a plant Heat Shock Protein (HSP)
XX family amino acid sequence (AAM51651-AAM51671). The transgenic plant has
XX increased stress tolerance, especially to heat. The plant is a cereal,
XX grass, ornamental plant, crop plant, food plant, oil-producing plant, a
XX synthetic product-producing plant, an environmental waste absorbing
XX plant, an alcohol plant, a medicinal plant, a recreational plant and/or
XX

```

---

```

CC an animal feed plant. In particular, the transgenic plant is cotton,
CC canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or
CC Arabidopsis thaliana. The plants may be used to produce animal feed,
CC alcohol, crop, oil, medicine or a synthetic product. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained from GenBank using the Accession Number reference provided
CC in the specification
XX Sequence 3105 BP; 891 A; 541 C; 852 G; 821 T; 0 U; 0 Other;
SQ Query Match 84.0%; Score 16.8; DB 4; Length 3105;
Best Local Similarity 90.0%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GAAACTGACCTTCACGCCTT 20
Db 1432 GAAATTGAACCTTCACGCCTT 1451

RESULT 11
AAI66066
ID AAI66066 standard; DNA; 6376 BP.
XX AAI66066;
XX 14-JAN-2002 (first entry)
XX Arabidopsis thaliana heat shock protein 101 gene (GenBank: AF218796).
XX Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
XX canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
XX Arabidopsis thaliana; ds.
XX Arabidopsis thaliana.
XX WO200170929-A2.
XX 27-SEP-2001.
XX 20-MAR-2001; 2001WO-US008836.
XX 20-MAR-2000; 2000US-0190769P.
XX 18-APR-2000; 2000US-0198116P.
XX (ARCH-) ARCH DEV CORP.
XX Lindquist S, Queitsch C, Vierling E;
XX WPI; 2001-639123/73.
XX P-PSDB; AAM51659.
XX Transgenic plants with improved heat stress tolerance, useful for
XX producing animal feed, oil and synthetic products.
XX Claim 4; Page; 91pp; English.
XX The invention relates to a transgenic plant, comprising a genetic
XX construct comprising a promoter operatively linked to a nucleic acid
XX sequence (AAI66057-AAI66084) encoding a plant Heat Shock Protein (HSP)
XX family amino acid sequence (AAM51651-AAM51671). The transgenic plant has
XX increased stress tolerance, especially to heat. The plant is a cereal,
XX grass, ornamental plant, crop plant, food plant, oil-producing plant, a
XX synthetic product-producing plant, an environmental waste absorbing
XX plant, an alcohol plant, a medicinal plant, a recreational plant and/or
XX

```



CC nucleotide polymorphisms are useful as a molecular genetic marker. These  
 CC sequences are useful for modifying a plant defense response, pest and/or  
 CC disease resistance, protein breakdown, glucan or chitin breakdown,  
 CC production of signals or elicitors for the manipulation of plant  
 CC development and/or host defense reaction in a plant and in gene therapy.  
 CC The protein sequences of the invention are useful for increasing plant  
 CC resistance to insect pest, increasing plant resistance to nematode  
 CC infection, reducing plant damage caused by herbivore predators, reducing  
 CC entry points for and spread of pathogens or reducing the reliance on  
 CC chemical application e.g., insecticides, nematocides or fungicides. The  
 CC present sequence is the Lolium perenne proteinase inhibitor f (LpPif)  
 CC consensus contig DNA.

SQ Sequence 596 BP; 153 A; 111 C; 153 G; 170 T; 0 U; 9 Other;

Query Match 80.0%; Score 16; DB 13; Length 596;  
 Best Local Similarity 100.0%; Pred. No. 4.2e+02;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 CTGACCTTCACGCGTT 20  
 |||||  
 Db 357 CTGACCTTCACGCGTT 342

RESULT 14  
 ADU07712  
 ID ADU07712 standard; DNA; 2751 BP.

XX AC ADU07712;

XX DT 13-JAN-2005 (first entry)

XX DE DNA sequence #213 encoding amylase.

XX KW Amylase; glucoamylase; 1,4-alpha-D-glucan glycohydrolase; alpha-amylase;  
 KW exoamylase; beta-amylase; glucosidic bond hydrolysis; starch; sugar;  
 KW glucose; maltodextrin; thermostable; gene; ds.

XX OS Unidentified.

XX PN WO2004091544-A2.

XX PD 28-OCT-2004.

XX PF 08-MAR-2004; 2004WO-US007096.

XX PR 06-MAR-2003; 2003US-00385305.

XX PR 28-MAR-2003; 2003US-0459014P.

XX PA (DIVE-) DIVERSA CORP.

XX PI Callen W, Richardson T, Frey G, Gray K, Kerovuo JS, Slupska M;  
 PI Barton N, O'donoghue E, Miller C;

XX WPI; 2004-775532/76.

XX PT Novel recombinant amylases, useful for hydrolysis of starch into sugars,  
 PT in detergent composition, for producing high-maltose or high-glucose  
 PT syrup, in brewing or alcohol production, bio-bleaching composition of  
 PT paper or pulp product.

XX PS Claim 4; SEQ ID NO 422; 436pp; English.

XX CC The invention relates to the isolation of polypeptides with amylase  
 CC activity, and polynucleotide sequences encoding the polypeptides. The  
 CC amylase activity comprises a glucoamylase activity, 1,4-alpha-D-glucan  
 CC glycohydrolase activity, alpha-amylase activity, exoamylase activity or  
 CC beta-amylase activity. The amylase activity comprises hydrolyzing  
 CC glucosidic bonds in a starch to produce sugars such as glucose and  
 CC maltodextrins. The amylase activity comprises cleaving a maltose or a D-  
 CC glucose unit from the non-reducing end of the starch. The amylase  
 CC activity is thermostable and the polypeptides of the invention are active  
 CC over a wide range of temperatures. The polypeptide sequences are useful

CC for making anti-amylase antibodies, for identifying a modulator of an  
 CC amylase activity, or identifying an amylase substrate. The polynucleotide  
 CC sequences are useful for producing recombinant polypeptides, and for  
 CC modifying codons in a nucleic acid encoding a polypeptide with an amylase  
 CC activity to increase or decrease its expression in a host cell. The  
 CC polypeptide and polynucleotide sequences are useful for making or  
 CC modifying a small molecule, and for determining a functional fragment of  
 CC an amylase enzyme. Primers to the polynucleotides are useful for  
 CC amplifying them. This sequence represents a polynucleotide sequence of  
 CC the invention.

SQ Sequence 2751 BP; 687 A; 838 C; 647 G; 579 T; 0 U; 0 Other;

Query Match 80.0%; Score 16; DB 13; Length 2751;  
 Best Local Similarity 100.0%; Pred. No. 5.2e+02;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACG 16

|||||

Db 2037 GAAACTGACCTTCACG 2052

RESULT 15

ACF85011

ID ACF85011 standard; DNA; 241 BP.

XX AC ACF85011;

XX DT 02-JUN-2005 (first entry)

XX DE Human SIRS/sepsis diagnostic marker DNA fragment 3871.

XX KW Systemic inflammatory response syndrome; SIRS; antibacterial;  
 KW immunosuppressive; antiinflammatory; diagnosis; sepsis; ds.

XX OS Homo sapiens.

XX PN WO2004087949-A2.

XX PD 14-OCT-2004.

XX PF 31-MAR-2004; 2004WO-EP003419.

XX PR 02-APR-2003; 2003DE-01015031.

XX PR 08-AUG-2003; 2003DE-01036511.

XX PR 02-SEP-2003; 2003DE-01040395.

XX PA (SIRS-) SIRS LAB GMBH.

XX PI Russwurm S, Reinhart K, Saluz H, Straube E, Zipfel PF, Deigner H;

XX WPI; 2004-748070/73.

XX PT In vitro detection of systemic inflammatory response syndrome and related  
 PT conditions, for e.g. monitoring progression, comprises detecting abnormal  
 PT expression of disease-related genes.

XX PS Disclosure; Page; 75pp; German.

XX CC The invention relates to a novel method for in vitro detection of  
 CC systemic inflammatory response syndrome (SIRS). The method comprises  
 CC detecting abnormal expression of disease-related genes, or their  
 CC associated peptides. The method of the invention demonstrates  
 CC antibacterial, immunosuppressive and antiinflammatory applications and  
 CC may be used for early differential diagnosis, monitoring progression,  
 CC assessing risk, assessing the likely response to treatment and for post  
 CC mortem diagnosis of systemic inflammatory response syndrome, sepsis and  
 CC sepsis-like conditions. The recombinant or synthetic nucleic acid  
 CC sequences of the invention, or derived proteins or peptides, may be  
 CC useful as calibrants in assays for the specified diseases, for evaluating  
 CC activity or toxicity in screening for active agents and/or for  
 CC preparation of agents for treatment or prevention of the specified  
 CC diseases. The current sequence is that of a human SIRS/sepsis diagnostic

CC marker DNA fragment of the invention. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at ftp.wipo.int/pub/published  
 CC pct sequences. Furthermore, a number of arbitrary SEQ ID NO.s are  
 CC disclosed within the specification, however, these have not been taken  
 CC into account during indexing due to inconsistencies in application and  
 CC format

XX SQ Sequence 241 BP; 66 A; 43 C; 63 G; 69 T; 0 U; 0 Other;

Query Match 79.0%; Score 15.8; DB 13; Length 241;

Best Local Similarity 89.5%; Pred. No. 4.7e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACGCCTT 19

DB 58 GAAACTGACCTTCATGCCT 76

RESULT 16

ABZ56651

ID ABZ56651 standard; cDNA; 299 BP.

XX AC ABZ56651;

XX DT 28-MAR-2003 (first entry)

XX DE Aspergillus oryzae polynucleotide SEQ ID NO 5764.

XX KW Aspergillus oryzae; fermentation; fungus; industrial; EST;

XX KW expressed sequence tag; gene; ss.

XX OS Aspergillus oryzae.

XX PN WO200279476-A1.

XX PD 10-OCT-2002.

XX PF 22-MAR-2002; 2002WO-IB000890.

XX PR 30-MAR-2001; 2001JP-00098371.

XX PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.

XX PA (NARE-) NAT RES INST BREWING.

XX PA (NORQ) NAT FOOD RES INST MIN AGRIC.

XX PI Machida M, Akita O, Kashiwagi Y, Kitamoto K, Horiuchi H;

XX PI Takeuchi M, Kobayashi T, Kitamoto N, Gomi K, Abe K;

XX WPI; 2003-046817/04.

XX PT Detection of expression of specific *Aspergillus* genes for monitoring the  
 XX fermentation and growth conditions of the fungus, using DNA probes.

XX PS Claim 1; SEQ ID NO 5764; 48pp + Sequence Listing; Japanese.

XX The invention relates to a polynucleotide having any of 6006 specific  
 CC sequences (ABZ50888-ABZ56893), which are expressed by a fungus under  
 CC specific culture conditions including one or more of eutrophic,  
 CC oligotrophic, solid, early germination, alkaline, high temperature, low  
 CC temperature or maltose culture or polynucleotides stringently hybridising  
 CC to these sequences. The polynucleotides are useful for monitoring the  
 CC progress of fermentation and the growth conditions of a fungus,  
 CC especially of *Aspergillus oryzae* which is widely used in industrial  
 CC fermentation. Also monitoring for fungal contamination. Note: The  
 CC sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 299 BP; 69 A; 80 C; 87 G; 63 T; 0 U; 0 Other;

Query Match 79.0%; Score 15.8; DB 8; Length 299;

Best Local Similarity 89.5%; Pred. No. 4.9e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAATGACCTTCACGCCTT 20

DB 171 ACACCGACCTTCACGCCTT 189

RESULT 17

ABZ56847

ID ABZ56847 standard; cDNA; 381 BP.

XX AC ABZ56847;

XX DT 28-MAR-2003 (first entry)

XX DE Aspergillus oryzae polynucleotide SEQ ID NO 5960.

XX KW Aspergillus oryzae; fermentation; fungus; industrial; EST;

XX KW expressed sequence tag; gene; ss.

XX OS Aspergillus oryzae.

XX PN WO200279476-A1.

XX PD 10-OCT-2002.

XX PF 22-MAR-2002; 2002WO-IB000890.

XX PR 30-MAR-2001; 2001JP-00098371.

XX PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.

XX PA (NARE-) NAT RES INST BREWING.

XX PA (NORQ) NAT FOOD RES INST MIN AGRIC.

XX PI Machida M, Akita O, Kashiwagi Y, Kitamoto K, Horiuchi H;

XX PI Takeuchi M, Kobayashi T, Kitamoto N, Gomi K, Abe K;

XX WPI; 2003-046817/04.

XX PT Detection of expression of specific *Aspergillus* genes for monitoring the  
 XX fermentation and growth conditions of the fungus, using DNA probes.

XX PS Claim 1; SEQ ID NO 5960; 48pp + Sequence Listing; Japanese.

XX The invention relates to a polynucleotide having any of 6006 specific  
 CC sequences (ABZ50888-ABZ56893), which are expressed by a fungus under  
 CC specific culture conditions including one or more of eutrophic,  
 CC oligotrophic, solid, early germination, alkaline, high temperature, low  
 CC temperature or maltose culture or polynucleotides stringently hybridising  
 CC to these sequences. The polynucleotides are useful for monitoring the  
 CC progress of fermentation and the growth conditions of a fungus,  
 CC especially of *Aspergillus oryzae* which is widely used in industrial  
 CC fermentation. Also monitoring for fungal contamination. Note: The  
 CC sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 381 BP; 76 A; 109 C; 110 G; 86 T; 0 U; 0 Other;

Query Match 79.0%; Score 15.8; DB 8; Length 381;

Best Local Similarity 89.5%; Pred. No. 5e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAATGACCTTCACGCCTT 20

DB 213 ACACCGACCTTCACGCCTT 231

RESULT 18

ABZ56844

ID ABZ56844 standard; cDNA; 382 BP.

XX AC ABZ56844;

```

XX 28-MAR-2003 (first entry)
XX Aspergillus oryzae polynucleotide SEQ ID NO 5957.
XX
XX Aspergillus oryzae; fermentation; fungus; industrial; EST;
XX expressed sequence tag; gene; ss.
XX
XX Aspergillus oryzae.
XX
XX WO200279476-A1.
XX
XX 10-OCT-2002.
XX
XX 22-MAR-2002; 2002WO-IB000890.
XX
XX 30-MAR-2001; 2001JP-00098371.
XX
XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
XX (NARE-) NAT RES INST BREWING.
XX (NORQ) NAT FOOD RES INST MIN AGRIC.
XX
XX Machida M, Akita O, Kashiwagi Y, Kitamoto K, Horiuchi H;
XX Takeuchi M, Kobayashi T, Kitamoto N, Gomi K, Abe K;
XX
XX WPI; 2003-046817/04.
XX
XX Detection of expression of specific Aspergillus genes for monitoring the
XX fermentation and growth conditions of the fungus, using DNA probes.
XX
XX Claim 1; SEQ ID NO 5957; 48pp + Sequence Listing; Japanese.
XX
XX The invention relates to a polynucleotide having any of 6006 specific
XX sequences (AB250888-AB256893), which are expressed by a fungus under
XX specific culture conditions including one or more of eutrophic,
XX oligotrophic, solid, early germination, alkaline, high temperature, low
XX temperature or maltose culture or polynucleotides stringently hybridising
XX to these sequences. The polynucleotides are useful for monitoring the
XX progress of fermentation and the growth conditions of a fungus,
XX especially of Aspergillus oryzae which is widely used in industrial
XX fermentation. Also monitoring for fungal contamination. Note: The
XX sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 382 BP; 77 A; 117 C; 99 G; 88 T; 0 U; 1 Other;
XX
XX Query Match 79.0%; Score 15.8; DB 8; Length 382;
XX Best Local Similarity 89.5%; Pred. No. 5e+02;
XX Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 2 AAACGTGACCTTCACGCCTT 20
XX ||| ||||| ||||| |||||
XX Db 70 ACACCGACCTTCACGCCTT 88
XX
XX RESULT 19
XX ID ABQ83840
XX AC ABQ83840;
XX
XX 29-JAN-2003 (first entry)
XX
XX Aspergillus oryzae nucleotide sequence SEQ ID NO:12.
XX
XX Aspergillus oryzae; fermentation; beer; wine; soy sauce; cooking sauce;
XX identification; fungal infection; gene; ss.
XX
XX Aspergillus oryzae.
XX
XX WO200279469-A1.
XX

```

```

PD 10-OCT-2002.
XX
XX 01-APR-2002; 2002WO-JP003267.
XX
XX 30-MAR-2001; 2001WO-JP002823.
XX
XX (TOHO-) TOHOKU TECHNOARCH KK.
XX
XX Abe K, Gomi K, Nakajima T, Yamagata Y, Hasegawa H, Iguchi Y;
XX
XX WPI; 2003-046815/04.
XX
XX Mold polynucleotide arrays of immobilized probes containing full or part
XX base sequences of mold-originated nucleic acids, for searching useful
XX molds for fermentation industry and identification of fungal infections.
XX
XX Example 3; Page 59; 62pp; Japanese.
XX
XX The present invention describes a method for obtaining polynucleotide
XX arrays, comprising immobilisation onto a support with a probe containing
XX the full or part base sequence of a mould-originated nucleic acid. Also
XX described: (1) detecting a target polynucleotide by labeling the target
XX polynucleotide for hybridisation with the probe immobilised in the array,
XX and detecting a signal from the hybridisation product; (2) identifying
XX closely-related moulds or their mutants by using the results obtained by
XX the detection method as indication; (3) screening useful and/or harmful
XX substances comprising isolation of nucleic acids from 1 or more moulds in
XX a candidate sample for labeling, hybridisation with the probe, detecting
XX a signal from the hybridisation product, and using the detection results
XX as indication; and (4) optimising expression of a target gene in a mould
XX by isolating a nucleic acid from the mould cultured under various
XX conditions for labeling, hybridisation with the probe, detecting a signal
XX from the hybridisation product, and using the detection results as
XX indication. The arrays are for searching useful moulds for fermentation
XX industry to make beer, wine, soy sauce and other cooking sauces, and
XX identification of fungal infections in human and cereals, making the
XX arrays important in the fermentation industry, medicine and agriculture.
XX The detection of moulds with these arrays can be easily achieved by
XX automation to save labour by carrying out hybridization of several
XX polynucleotides simultaneously and globally, including the identification
XX of mutants and optimisation of fermentation process, and screening useful
XX and harmful substances. The present sequence represents an Aspergillus
XX oryzae nucleotide sequence which is used in an example from the present
XX invention
XX
XX Sequence 440 BP; 96 A; 132 C; 112 G; 100 T; 0 U; 0 Other;
XX
XX Query Match 79.0%; Score 15.8; DB 8; Length 440;
XX Best Local Similarity 89.5%; Pred. No. 5.1e+02;
XX Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 2 AAACGTGACCTTCACGCCTT 20
XX ||| ||||| ||||| |||||
XX Db 204 ACACCGACCTTCACGCCTT 222
XX
XX RESULT 20
XX ABV77524
XX ID ABV77524 standard; DNA; 440 BP.
XX
XX AC ABV77524;
XX
XX 04-FEB-2003 (first entry)
XX
XX A. oryzae array probe sequence #12.
XX
XX Polynucleotide array; fungus; fermentation; beer; wine; soy sauce;
XX fungal infection; probe; ss.
XX
XX Aspergillus oryzae.
XX
XX WO200281701-A1.
XX

```



```
PD 17-OCT-2002.
XX
XX 30-MAR-2001; 2001WO-JP002823.
XX
XX 30-MAR-2001; 2001WO-JP002823.
XX
XX (TOHO-) TOHOKU TECHNOARCH KK.
XX
XX Abe K, Gomi K, Nakajima T, Yamagata Y, Hasegawa H, Iguchi Y;
XX WPI; 2003-092938/08.
XX
XX Fungal polynucleotide arrays of immobilized probes containing full or
XX part base sequences of fungus-originated nucleic acids, for searching
XX useful fungi for fermentation industry and identification of fungal
XX infections.
XX
XX Disclosure; Page 38; 43pp; Japanese.
XX
XX The invention relates to a novel polynucleotide arrays, obtained by
XX immobilisation onto a support with a probe containing the full or part
XX base sequence of a fungus-originated nucleic acid. The arrays are useful
XX for searching useful fungi for fermentation industry to make beer, wine,
XX soy sauce and other cooking sauces, and identification of fungal
XX infections in human and cereals, making such arrays important in the
XX fermentation industry, medicine and agriculture. The detection of fungi
XX with these arrays can be easily achieved by automation to save labour by
XX carrying out hybridisation of plural polynucleotides simultaneously and
XX globally, including the identification of mutants and optimisation of
XX fermentation process, and screening useful and harmful substances. The
XX sequence represents a polynucleotide sequence from A. oryzae used in an
XX array of the invention
XX
XX Sequence 440 BP; 96 A; 132 C; 112 G; 100 T; 0 U; 0 Other;
XX
XX Query Match 79.0%; Score 15.8; DB 8; Length 440;
XX Best Local Similarity 89.5%; Pred. No. 5.1e+02;
XX Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 2 AAACGTGACCTTCACGCGCTT 20
XX |||||
XX 204 ACACCGACCTTCACGCGCTT 222
XX
XX RESULT 21
XX ABZ56877
XX ID ABZ56877 standard; cDNA; 497 BP.
XX
XX AC ABZ56877;
XX
XX DT 28-MAR-2003 (first entry)
XX
XX DE Aspergillus oryzae polynucleotide SEQ ID NO 5990.
XX
XX KW Aspergillus oryzae; fermentation; fungus; industrial; EST;
XX expressed sequence tag; gene; ss.
XX
XX OS Aspergillus oryzae.
XX
XX PN WO200279476-A1.
XX
XX 10-OCT-2002.
XX
XX 22-MAR-2002; 2002WO-IB000890.
XX
XX 30-MAR-2001; 2001JP-00098371.
XX
XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
XX (NARE-) NAT RES INST BREWING.
XX (NORQ) NAT FOOD RES INST MIN AGRIC.
XX
XX Machida M, Akita O, Kashiwagi Y, Kitamoto K, Horiuchi H;
XX Takeuchi M, Kobayashi T, Kitamoto N, Gomi K, Abe K;
XX
```

---

```
XX WPI; 2003-046817/04.
XX
XX Detection of expression of specific Aspergillus genes for monitoring the
XX fermentation and growth conditions of the fungus, using DNA probes.
XX
XX Claim 1; SEQ ID NO 5990; 48pp + Sequence Listing; Japanese.
XX
XX The invention relates to a polynucleotide having any of 6006 specific
XX sequences (ABZ50888-ABZ56893), which are expressed by a fungus under
XX specific culture conditions including one or more of eutrophic,
XX oligotrophic, solid, early germination, alkaline, high temperature, low
XX temperature or maltose culture or polynucleotides stringently hybridising
XX to these sequences. The polynucleotides are useful for monitoring the
XX progress of fermentation and the growth conditions of a fungus,
XX especially of Aspergillus oryzae which is widely used in industrial
XX fermentation. Also monitoring for fungal contamination. Note: The
XX sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 497 BP; 102 A; 146 C; 136 G; 111 T; 0 U; 2 Other;
XX
XX Query Match 79.0%; Score 15.8; DB 8; Length 497;
XX Best Local Similarity 89.5%; Pred. No. 5.2e+02;
XX Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 2 AAACGTGACCTTCACGCGCTT 20
XX |||||
XX 174 ACACCGACCTTCACGCGCTT 192
XX
XX Db
XX
XX RESULT 22
XX ADS31337
XX ID ADS31337 standard; DNA; 510 BP.
XX
XX AC ADS31337;
XX
XX DT 18-NOV-2004 (first entry)
XX
XX DE Human genome high complexity repeat found in the HIRA gene #370.
XX
XX OS Homo sapiens.
XX
XX PN US2003224356-A1.
XX
XX PD 04-DEC-2003.
XX
XX PF 14-MAY-2001; 2001US-00854867.
XX
XX PR 16-MAY-2000; 2000US-00573080.
XX
XX (KNOL/) KNOLL J H M.
XX (ROGA/) ROGAN P K.
XX
XX Knoll JHM, Rogan PK;
XX
XX WPI; 2002-062378/08.
XX
XX Single copy genomic hybridization probes for detecting specific nucleic
XX acid sequences in sample by in situ hybridization useful for detection of
XX acquired or inherited genetic diseases.
XX
XX Example 1; SEQ ID NO 370; 30pp; English.
XX
XX The invention relates to a nucleic acid hybridisation probe comprising a
XX labelled, single copy nucleic acids of at least 50 nucleotides, which
```

CC will hybridise to a deduced single copy sequence interval in target  
CC nucleic acid (TNA) of known sequence. The single copy sequence is deduced  
CC by comparing the target nucleic acid (e.g. a disease causing gene) with a  
CC collection of high and low complexity repeat sequences as found in the  
CC genome of the organism from containing the target nucleic acid. The probe  
CC is generated by PCR on the target sequence. The probe is essentially free  
CC of blocking nucleic acid sequences which will hybridise to repeat  
CC sequences within the genome of which the TNA is a part, and is labelled  
CC with a label selected from fluorochrome-responsive labels, fluorochromes,  
CC calorimetric chemical, conjugated proteins, antibodies, antigens and  
CC their mixtures. The probe is useful in a hybridisation method, where the  
CC hybridisation method is from in situ hybridisation, Southern blot, and  
CC other methods in which nucleic acid is immobilised, where the method  
CC further comprises selecting a single copy nucleic acid which will  
CC hybridise to a duplication or triplicon sequence domain. The probe is useful  
CC for determining the existence of previously unknown repeat sequence  
CC families in a genome. The method comprises reacting a labelled probe with  
CC the genome, causing the probe to hybridise and ascertaining if the probe  
CC hybridises to the genome at more than three preferably ten different  
CC locations as a determination of new repeat sequence family, where the  
CC determining step comprises selecting the single copy sequence from a  
CC duplication or triplicon sequence domain. The probe is useful for  
CC determining a chromosome breakpoint and is useful in the fields for  
CC cytogenetics and molecular genetics for determining the presence of  
CC specific nucleic acid sequences in a sample of eukaryotic origin, e.g.  
CC situ hybridisation as a detection of acquired or inherited genetic  
CC diseases especially for detection of genetic or neoplastic disorders.  
CC Unlike prior art techniques, the probe permits more precise chromosomal  
CC breakpoint determinations by in situ hybridisation. The genomic sequence  
CC comprising the human HIRA gene (histone cell cycle regulation defective,  
CC S. cerevisiae, homologue A) was analysed for single copy sequence  
CC intervals for use as probes of the invention. HIRA is located on  
CC chromosome 22 as a duplicate, deletions of 1 copy lead to DiGeorge and  
CC Velo-Cardio-facial syndromes. The present sequence is a high complexity  
CC repeat found within the human genome used to analyse the HIRA gene for  
CC repeat regions. Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from USPTO at [seqdata.uspto.gov/sequence.html?DocID=20030224356](http://seqdata.uspto.gov/sequence.html?DocID=20030224356).

XX Sequence 510 BP; 114 A; 154 C; 89 G; 153 T; 0 U; 0 Other;

Query Match 79.0%; Score 15.8; DB 7; Length 510;

Best Local Similarity 89.5%; Pred. No. 5.2e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAACCTGACCTTCACGCCTT 20

||||| ||||| ||||| |||||

Db 217 AAACCTGATCTCCCGCCTT 235

RESULT 23

ADY36725

ID ADY36725 standard; DNA; 510 BP.

XX AC ADY36725;

XX DT 05-MAY-2005 (first entry)

XX DE HIRA genomic fragment SEQ ID NO 370.

XX KW hybridization; DNA detection; neoplasm; genetic disorder; cytogenetics;

XX KW HIRA; db.

OS Homo sapiens.

XX PN WO200186089-A2.

XX XX 22-NOV-2001.

XX PF 15-MAY-2001; 2001WO-US015674.

XX XX 16-MAY-2000; 2000US-00573080.

PR 14-MAY-2001; 2001US-00854867.

XX (CHIL-) CHILDREN'S MERCY HOSPITAL.

FA Knoll JHM, Rogan PK, Casarro PM;

PI WPI; 2002-062378/08.

XX Single copy genomic hybridization probes for detecting specific nucleic

PT acid sequences in sample by in situ hybridization useful for detection of

PT acquired or inherited genetic diseases.

PS Example 1; SEQ ID NO 370; 67pp; English.

XX The invention describes a nucleic acid hybridization probe (I) comprising

CC a labeled, single copy nucleic acid of at least 50 nucleotides, which

CC will hybridize to a deduced single copy sequence interval in target

CC nucleic acid (TNA) of known sequence. (I) is useful in a hybridization

CC method which comprises preparing a reaction mixture comprising TNA and

CC (I) which hybridizes to TNA, and causing (I) to hybridize to TNA, where

CC the hybridization method is from in situ hybridization, Southern blot,

CC and other methods in which nucleic acid is immobilized, where the method

CC further comprises selecting a single copy nucleic acid which will

CC hybridize to a duplication or triplicon sequence domain. (I) is useful for:

CC determining the existence of previously unknown repeat sequence families

CC in a genome; determining a chromosome breakpoint and in the fields of

CC cytogenetics and molecular genetics for determining the presence of

CC specific nucleic acid sequences in a sample of eukaryotic origin, e.g.

CC the probes may be used to analyze specific chromosomal locations by in

CC situ hybridization as a detection of acquired or inherited genetic

CC diseases especially for detection of genetic or neoplastic disorders.

CC Unlike prior art techniques, (I) permits more precise chromosomal

CC breakpoint determinations by in situ hybridization. Hybridization

CC techniques utilizing (I), have made it possible to obtain reliable,

CC easily detectable signals with relatively small probes. A readily

CC detectable signal was obtained with a probe on the order of 2 kb in

CC length, using fluorescent in situ hybridization (FISH) technology. This

CC sensitivity of (I) is improved compared to the prior art, because the

CC probes of (I) are homogeneous single copy sequences. However, smaller

CC amplified segments, each comprising non-repetitive sequences, may also be

CC used in combination as probes to achieve adequate signals for in situ

CC hybridization. Complex single copy probes that hybridize to duplicated or

CC triplicated targets can also increase hybridization signals. This

CC sequence represents a human HIRA genomic sequence that shows homology to

CC a known high-complexity repeat sequence family of the human genome and is

CC used in the creation of an HIRA gene probe.

XX Sequence 510 BP; 114 A; 154 C; 89 G; 153 T; 0 U; 0 Other;

Query Match 79.0%; Score 15.8; DB 7; Length 510;

Best Local Similarity 89.5%; Pred. No. 5.2e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAACCTGACCTTCACGCCTT 20

||||| ||||| ||||| |||||

Db 217 AAACCTGATCTCCCGCCTT 235

RESULT 24

ABZ56884

ID ABZ56884 standard; cDNA; 522 BP.

XX AC ABZ56884;

XX DT 28-MAR-2003 (first entry)

XX DE Aspergillus oryzae polynucleotide SEQ ID NO 5997.

XX XX Aspergillus oryzae; fermentation; fungus; industrial; EST;

XX KW expressed sequence tag; gene; ss.

XX OS Aspergillus oryzae.

```
PN WO200279476-A1.
XX
PD 10-OCT-2002.
XX
PF 22-MAR-2002; 2002WO-IB0000890.
XX
XX 30-MAR-2001; 2001JP-00098371.
XX
XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA (NARE-) NAT RES INST BREWING.
PA (NORQ) NAT FOOD RES INST MIN AGRIC.
XX
PI Machida M, Akita O, Kashiwagi Y, Kitamoto K, Horiuchi H;
PI Takeuchi M, Kobayashi T, Kitamoto N, Gomi K, Abe K;
XX
XX WPI; 2003-046817/04.
XX
PT Detection of expression of specific Aspergillus genes for monitoring the
PT fermentation and growth conditions of the fungus, using DNA probes.
XX
XX Claim 1; SEQ ID NO 5997; 48pp + Sequence Listing; Japanese.
XX
CC The invention relates to a polynucleotide having any of 6006 specific
CC sequences (ABZ50888-ABZ56893), which are expressed by a fungus under
CC specific culture conditions including one or more of autotrophic,
CC oligotrophic, solid, early germination, alkaline, high temperature, low
CC temperature or maltose culture or polynucleotides stringently hybridising
CC to these sequences. The polynucleotides are useful for monitoring the
CC progress of fermentation and the growth conditions of a fungus,
CC especially of Aspergillus oryzae which is widely used in industrial
CC fermentation. Also monitoring for fungal contamination. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 522 BP; 111 A; 164 C; 117 G; 125 T; 0 U; 5 Other;

Query Match 79.0%; Score 15.8; DB 8; Length 522;
Best Local Similarity 89.5%; Pred. No. 5.3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AACTGACCTTCACGCCTT 20
| | | | | | | | | | | | | | | |
DB 264 ACACCGACCTTCACGCCTT 282

RESULT 25
AAF08072
ID AAF08072 standard; cDNA; 641 BP.
XX
AC AAF08072;
XX
DT 13-MAR-2001 (first entry)
XX
DE Fusarium venenatum EST SEQ ID NO:595.
XX
KW Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS Fusarium venenatum.
XX
XX WO200056762-A2.
PN
XX 28-SEP-2000.
PD
XX
PF 22-MAR-2000; 2000WO-US007781.
XX
XX 22-MAR-1999; 99US-00273623.
PR
XX (NOVO) NOVO NORDISK BIOTECH INC.
PA

(PA) (NOVO) NOVO NORDISK AS.
XX
PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX
XX WPI; 2000-594572/56.
XX
XX Monitoring differential expression of genes in filamentous fungal cells
XX uses fluorescence-labeled nucleic acids isolated from the cells and a
XX substrate of expressed sequence tags.
XX
XX Claim 86; Page 613; 3161pp; English.
XX
XX The present invention describes a method for monitoring differential
XX expression of genes in a first filamentous fungal (FF) cell relative to
XX expression of the same genes in one or more second filamentous fungal
XX cells. The method uses fluorescence-labeled nucleic acids isolated from
XX the FF cells and a substrate of expressed sequence tags (EST). The ESTs
XX are used in the methods for monitoring differential expression of genes
XX in a first filamentous fungal (FF) cell relative to expression of the
XX same genes in one or more second filamentous fungal cells. Monitoring the
XX global expression of genes from FF cells allows the production potential
XX of the microorganisms to be improved. New genes may be discovered.
XX possible functions of unknown open reading frames can be identified and
XX gene copy number variation and stability can be monitored. The expression
XX of genes can be used to study how FF cells adapt to changes in culture
XX conditions, environmental stress, spore morphogenesis, recombination,
XX metabolic or catabolic pathway engineering. Using ESTs provides several
XX advantages over genomic or random cDNA clones including elimination of
XX redundancy as one spot on an array equals one gene or open reading frame,
XX and organisation of the microarrays based on function of the gene
XX products to facilitate analysis of the results. AAF07478 to AAF11247
XX represents ESTs from Fusarium venenatum; AAF11248 to AAF11853 represents
XX ESTs from Aspergillus niger; AAF11854 to AAF14878 represents ESTs from
XX Aspergillus oryzae; and AAF14879 to AAF15337 represents ESTs from
XX Trichoderma reesei, which are all specifically claimed in the present
XX invention
XX
SQ Sequence 641 BP; 145 A; 173 C; 150 G; 164 T; 0 U; 9 Other;

Query Match 79.0%; Score 15.8; DB 3; Length 641;
Best Local Similarity 89.5%; Pred. No. 5.4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACGCCTT 19
| | | | | | | | | | | | | | | |
DB 469 GAAACTGCTTCACGACT 487

RESULT 26
ADU52113
ID ADU52113 standard; cDNA; 641 BP.
XX
AC ADU52113;
XX
DT 10-FEB-2005 (first entry)
XX
DE Fusarium venenatum CC1-3 EST, SEQ ID NO:595.
XX
KW Gene expression; biochip; microarray; hybridization; EST;
KW expressed sequence tag; ss.
XX
OS Fusarium venenatum; CC1-3.
XX
XX US2004229367-A1.
PN
XX 18-NOV-2004.
PD
XX
PF 29-AUG-2003; 2003US-00653047.
XX
XX 22-MAR-1999; 99US-00273623.
PR
XX 22-MAR-2000; 2000US-00533559.
XX
XX (NOVO) NOVOZYMES BIOTECH INC.
PA
```

```
PA (NOVO ) NOVOZYMES INC AS.
XX
XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX WPI; 2004-832481/82.
XX
XX Monitoring differential expression of genes in a filamentous fungal cell
XX relative to expression of the genes in second filamentous fungal cells,
XX comprises using microarrays containing Fusarium venenatum expressed
XX sequence tags.
XX
XX Claim 1; SEQ ID NO 595; 274pp; English.
XX
XX The invention relates to a method for monitoring the differential
XX expression of genes in a first filamentous fungal cell relative to the
XX same genes in one or more second filamentous fungal cells. The method
XX involves differentially labeling nucleic acids from the first and second
XX fungal cells with fluorescent reporter dyes, and adding the mixture of
XX fluorescently labeled polynucleotides to a substrate containing an array
XX of Fusarium venenatum expressed sequence tags (ESTs) selected from
XX ADU51519-ADU5288 under conditions that permit hybridization. Relative
XX expression of genes in the fungal cells is determined by fluorescence;
XX the fluorescent signal emitted when labeled nucleic acids from the first
XX fungal cell bind to the ESTs in the array differs in color from that
XX emitted when labeled nucleic acids from the second fungal cell hybridise.
XX When both sets of labeled nucleic acids bind to the array, a distinct
XX combined fluorescence emission color is produced. The filamentous fungal
XX cells used in the method are chosen from Acremonium, Aspergillus,
XX Fusarium, Humicola, Mucor, Myceliophthora, Neurospora, Penicillium,
XX Thielavia, Tolypocladium, and Trichoderma cells, preferably Fusarium
XX venenatum, Aspergillus niger, or Aspergillus oryzae. The method of the
XX invention is useful for monitoring the expression of a plurality of genes
XX in filamentous fungal cells, in order to improve the cells' protein
XX production capacity when such organisms are used for the industrial
XX production of proteins (e.g., enzymes). The present sequence represents
XX an EST derived from Fusarium venenatum CC1-3 (a morphological mutant of
XX strain ATCC 20334) which is specifically claimed for use in the method of
XX the invention. Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic format
XX directly from the US patent office at
XX seqdata.uspto.gov/sequence.html?DocID=US20040222367.
XX
XX Sequence 641 BP; 145 A; 173 C; 150 G; 164 T; 0 U; 9 Other;
XX
XX Query Match 79.0%; Score 15.8; DB 13; Length 641;
XX Best Local Similarity 89.5%; Pred. No. 5.4e+02;
XX Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 1 GAAACTGACCTTCACGCCT 19
XX ||||| ||||| ||||| ||||| |||||
XX Db 469 GAAACTGCTCTCAGCACT 487
XX
XX RESULT 27
XX AD290116
XX ID AD290116 standard; cDNA; 641 BP.
XX
XX AC AD290116;
XX
XX DT 11-AUG-2005 (first entry)
XX
XX DE Fusarium venenatum expressed sequence tag cDNA SEQ ID NO 595.
XX
XX KW differential expression; gene expression; filamentous fungus;
XX expressed sequence tag; EST; variation; microarray; ss.
XX
XX OS Fusarium venenatum.
XX
XX PN US6902887-B1.
XX
XX PD 07-JUN-2005.
XX
XX PP 22-MAR-2000; 2000US-00533559.
XX
XX 22-MAR-1999; 99US-00273623.
XX
XX (NOVO ) NOVOZYMES BIOTECH INC.
XX (NOVO ) NOVOZYMES AS.
XX
XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX WPI; 2005-401635/41.
XX
XX Monitoring differential expression of genes in first filamentous fungal
XX cell relative to that of genes in second fungal cells, comprises
XX hybridizing fluorescence labeled nucleic acids from the cells to an array
XX of expressed sequence tags.
XX
XX Disclosure; SEQ ID NO 595; 264pp; English.
XX
XX The invention relates to a method of monitoring (M1) differential
XX expression of multiple genes in a first filamentous fungal cell relative
XX to expression of the same genes in one or more second filamentous fungal
XX cells, comprises adding fluorescence labeled nucleic acids isolated from
XX the cells to a substrate containing an array of Aspergillus oryzae
XX expressed sequence tag (EST) of AD293898-AD296922, and examining the
XX array under fluorescence excitation conditions. (M1) is useful for
XX monitoring global expression of several genes from a filamentous fungal
XX cell, discovering new genes, identifying possible functions of unknown
XX open reading frames and monitoring gene copy number variation and
XX stability. In (M1), one spot on an array equals one gene or open reading
XX frame, extensive follow-up characterization is unnecessary since sequence
XX information is available and EST microarrays can be organized based on
XX function of the gene products. This sequence corresponds to an EST
XX sequence of the invention. (Note: this sequence is not given in the
XX printed specification but can be obtained in electronic form from the
XX USPTO web site seqdata.uspto.gov/sequence.html; Document ID: 6902887B1).
XX
XX Sequence 641 BP; 145 A; 173 C; 150 G; 164 T; 0 U; 9 Other;
XX
XX Query Match 79.0%; Score 15.8; DB 14; Length 641;
XX Best Local Similarity 89.5%; Pred. No. 5.4e+02;
XX Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 1 GAAACTGACCTTCACGCCT 19
XX ||||| ||||| ||||| ||||| |||||
XX Db 469 GAAACTGCTCTCAGCACT 487
XX
XX RESULT 28
XX ACF68857
XX ID ACF68857 standard; DNA; 1173 BP.
XX
XX AC ACF68857;
XX
XX DT 20-NOV-2003 (first entry)
XX
XX DE Photorhabdus luminescens nucleotide sequence #7324.
XX
XX KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
XX detection; food; gene expression; plant; animal; microorganism; toxin;
XX antibiotic; biopesticide; virulence factor; disease model; plague;
XX whooping cough; gene; ds.
XX
XX OS Photorhabdus luminescens.
XX
XX PN WO200294867-A2.
XX
XX PD 28-NOV-2002.
XX
XX PF 07-FEB-2002; 2002WO-IB003040.
XX
XX PR 07-FEB-2001; 2001FR-00001659.
XX
XX (INSP ) INST PASTEUR.
XX (CNRS ) CNRS CENT NAT RECH SCI.
```

XX Duchaud E, Taourit S, Glaeser P, Frangeul L, Kunst P, Danchin A;  
 PI Buchrieser C;  
 XX WPI; 2003-148459/14.  
 XX Genomic sequence of *Photobacterium luminescens* and encoded polypeptides,  
 PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.  
 XX  
 XX Claim 2; SEQ ID NO 7324; 1205pp; French.  
 XX  
 XX The invention relates to the isolation of genes and their encoded  
 CC proteins from *Photobacterium luminescens*. The isolated sequences are  
 CC sources of probes and primers for detecting the genome of *P. luminescens*  
 CC and related species; to study polymorphisms; for gene analysis and for  
 CC detection/amplification of the genes. Antibodies (Ab) raised against the  
 CC polypeptides encoded by the genes are used for detection/identification  
 CC of *P. luminescens*, e.g. in foods. The genes, proteins, Ab and cells that  
 CC carry a gene-containing vector are used to select compounds that  
 CC modulate, regulate, induce or inhibit expression of the genes in plants,  
 CC animals or microorganisms other than *P. luminescens* and are able to alter  
 CC response or sensitivity to toxins and antibiotics produced by *P.*  
 CC *luminescens*. Cells transformed to express the genes are useful for  
 CC recombinant production of the proteins, particularly toxins and  
 CC antibacterials useful as insecticides, bactericides and fungicides. The  
 CC genes, proteins, vectors containing the genes and Ab are also useful  
 CC therapeutically (to treat microbial infection by bacteria or fungi that  
 CC are sensitive to *P. luminescens*-encoded toxins or antibiotics) and as  
 CC biopesticides. Other uses of the genes and the proteins are as virulence  
 CC factors and for identifying targets of human diseases for which *P.*  
 CC *luminescens* is a model (particularly plague and whooping cough). This  
 CC sequence represents one of the isolated *P. luminescens* genes  
 XX  
 XX Sequence 1173 BP; 233 A; 232 C; 333 G; 375 T; 0 U; 0 Other;  
 SQ

Query Match 79.0%; Score 15.8; DB 10; Length 1173;  
 Best Local Similarity 89.5%; Pred. No. 5.9e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAACTGACCTTCAGCGCT 19  
 |||||  
 Db 624 GAACTGACCTTCAGCGCT 642  
 |||||

RESULT 29  
 AAF13166  
 ID AAF13166 standard; cDNA; 1362 BP.  
 XX  
 AC AAF13166;  
 XX  
 DT 13-MAR-2001 (first entry)  
 XX  
 DE *Aspergillus oryzae* EST SEQ ID NO:5689.  
 XX  
 KW Multiple gene expression; filamentous fungal cell; EST;  
 KW expressed sequence tag; *Fusarium venenatum*; *Aspergillus niger*;  
 KW *Aspergillus oryzae*; *Trichoderma reesei*; identification; recombination;  
 KW culture condition; environmental stress; spore morphogenesis;  
 KW metabolic pathway engineering; catabolic pathway engineering; ss.  
 XX  
 OS *Aspergillus oryzae*.  
 XX  
 PN WO200056762-A2.  
 XX  
 PD 28-SEP-2000.  
 XX  
 XX 22-MAR-2000; 2000WO-US007781.  
 PF  
 XX 22-MAR-1999; 99US-00273623.  
 PR  
 XX (NOVO ) NOVO NORDISK BIOTECH INC.  
 PA (NOVO ) NOVO NORDISK AS.  
 XX

PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;  
 XX WPI; 2000-594572/56.  
 XX  
 XX Monitoring differential expression of genes in filamentous fungal cells  
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a  
 PT substrate of expressed sequence tags.  
 XX  
 XX Claim 88; Page 2357-2358; 3161pp; English.  
 XX  
 XX The present invention describes a method for monitoring differential  
 CC expression of genes in a first filamentous fungal (FF) cell relative to  
 CC expression of the same genes in one or more second filamentous fungal  
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from  
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs  
 CC are used in the methods for monitoring differential expression of genes  
 CC in a first filamentous fungal (FF) cell relative to expression of the  
 CC same genes in one or more second filamentous fungal cells. Monitoring the  
 CC global expression of genes from FF cells allows the production potential  
 CC of the microorganisms to be improved. New genes may be discovered,  
 CC possible functions of unknown open reading frames can be identified and  
 CC gene copy number variation and stability can be monitored. The expression  
 CC of genes can be used to study how FF cells adapt to changes in culture  
 CC conditions, environmental stress, spore morphogenesis, recombination,  
 CC metabolic or catabolic pathway engineering. Using ESTs provides several  
 CC advantages over genomic or random cDNA clones including elimination of  
 CC redundancy as one spot on an array equals one gene or open reading frame,  
 CC and organization of the microarrays based on function of the gene  
 CC products to facilitate analysis of the results. AAF07478 to AAF11247  
 CC represents ESTs from *Fusarium venenatum*; AAF11248 to AAF11853 represents  
 CC ESTs from *Aspergillus niger*; AAF11854 to AAF14878 represents ESTs from  
 CC *Aspergillus oryzae*; and AAF14879 to AAF15337 represents ESTs from  
 CC *Trichoderma reesei*, which are all specifically claimed in the present  
 CC invention  
 XX  
 XX Sequence 1362 BP; 299 A; 396 C; 348 G; 319 T; 0 U; 0 Other;  
 SQ

Query Match 79.0%; Score 15.8; DB 3; Length 1362;  
 Best Local Similarity 89.5%; Pred. No. 6e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAATGACCTTCAGCGCTT 20  
 |||||  
 Db 331 ACACCGACCTTCAGCGCTT 349  
 |||||

RESULT 30  
 ADU57207  
 ID ADU57207 standard; cDNA; 1362 BP.  
 XX  
 AC ADU57207;  
 XX  
 DT 10-FEB-2005 (first entry)  
 XX  
 DE *Aspergillus oryzae* strain strain A1560/strain AL-1 EST, SEQ ID NO:5689.  
 XX  
 KW Gene expression; biochip; microarray; hybridization; EST;  
 KW expressed sequence tag; ss.  
 XX  
 OS *Aspergillus oryzae*.  
 XX  
 PN US2004229367-A1.  
 XX  
 PD 18-NOV-2004.  
 XX  
 XX 29-AUG-2003; 2003US-00653047.  
 PF  
 XX 22-MAR-1999; 99US-00273623.  
 PR  
 XX 22-MAR-2000; 2000US-00533559.  
 PR  
 XX (NOVO ) NOVOZYMES BIOTECH INC.  
 PA (NOVO ) NOVOZYMES INC AS.  
 XX

PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olesen PB;  
 XX WPI; 2004-832481/82.  
 XX  
 PT Monitoring differential expression of genes in a filamentous fungal cell  
 PT relative to expression of the genes in second filamentous fungal cells,  
 PT comprises using microarrays containing *Fusarium venenatum* expressed  
 PT sequence tags.  
 XX  
 XX Example 12; SEQ ID NO 5689; 274pp; English.  
 PS  
 XX The invention relates to a method for monitoring the differential  
 CC expression of genes in a first filamentous fungal cell relative to the  
 CC same genes in one or more second filamentous fungal cells. The method  
 CC involves differentially labeling nucleic acids from the first and second  
 CC fungal cells with fluorescent reporter dyes, and adding the mixture of  
 CC fluorescently labeled polynucleotides to a substrate containing an array  
 CC of *Fusarium venenatum* expressed sequence tags (ESTs) selected from  
 CC ADU51519-ADU55288 under conditions that permit hybridization. Relative  
 CC expression of genes in the fungal cells is determined by fluorescence;  
 CC the fluorescent signal emitted when labeled nucleic acids from the first  
 CC fungal cell bind to the ESTs in the array differs in color from that  
 CC emitted when labeled nucleic acids from the second fungal cell hybridize.  
 CC When both sets of labeled nucleic acids bind to the array, a distinct  
 CC combined fluorescence emission color is produced. The filamentous fungal  
 CC cells used in the method are chosen from *Acremonium*, *Aspergillus*,  
 CC *Fusarium*, *Humicola*, *Mucor*, *Myceliophthora*, *Neurospora*, *Penicillium*,  
 CC *Thielavia*, *Tolypocladium*, and *Trichoderma* cells, preferably *Fusarium*  
 CC *venenatum*, *Aspergillus niger*, or *Aspergillus oryzae*. The method of the  
 CC invention is useful for monitoring the expression of a plurality of genes  
 CC in filamentous fungal cells, in order to improve the cells' protein  
 CC production capacity when such organisms are used for the industrial  
 CC production of proteins (e.g., enzymes). Sequences ADU5895-ADU5919  
 CC represent ESTs derived from *Aspergillus oryzae* strain AI560 and/or strain  
 CC AL-1 which were obtained in an example of the invention. Note: The  
 CC sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the US  
 CC patent office at seqdata.uspto.gov/sequence.html?DocID=US20040222367.  
 XX  
 SQ Sequence 1362 BP; 299 A; 396 C; 348 G; 319 T; 0 U; 0 Other;  
 Query Match 79.0%; Score 15.8; DB 13; Length 1362;  
 Best Local Similarity 89.5%; Pred. No. 6e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 AAACGTGACCTTCACGCCTT 20  
 DB 331 ACACCGACCTTCACGCCTT 349  
 ||| ||||||||||||  
 RESULT 31  
 ADZ95210  
 ID ADZ95210 standard; cDNA; 1362 BP.  
 XX  
 AC ADZ95210;  
 XX  
 DT 11-AUG-2005 (first entry)  
 XX  
 DE *Aspergillus oryzae* expressed sequence tag cDNA SEQ ID NO 5689.  
 XX differential expression; gene expression; filamentous fungus;  
 KW expressed sequence tag; EST; variation; microarray; ss.  
 XX  
 OS *Aspergillus oryzae*.  
 XX  
 PN US6902887-B1.  
 XX  
 XX 07-JUN-2005.  
 PD  
 XX 22-MAR-2000; 2000US-00533559.  
 PF  
 XX 22-MAR-1999; 99US-00273623.  
 PR  
 XX

PA (NOVO ) NOVOZYMES BIOTECH INC.  
 XX (NOVO ) NOVOZYMES AS.  
 PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olesen PB;  
 XX WPI; 2005-401635/41.  
 DR  
 XX Monitoring differential expression of genes in first filamentous fungal  
 PT cell relative to that of genes in second fungal cells, comprises  
 PT hybridizing fluorescence labeled nucleic acids from the cells to an array  
 PT of expressed sequence tags.  
 XX  
 XX Claim 1; SEQ ID NO 5689; 264pp; English.  
 PS  
 XX The invention relates to a method of monitoring (M1) differential  
 CC expression of multiple genes in a first filamentous fungal cell relative  
 CC to expression of the same genes in one or more second filamentous fungal  
 CC cells, comprises adding fluorescence labeled nucleic acids isolated from  
 CC the cells to a substrate containing an array of *Aspergillus oryzae*  
 CC expressed sequence tag (EST) of ADZ93898-ADZ96922, and examining the  
 CC array under fluorescence excitation conditions. (M1) is useful for  
 CC monitoring global expression of several genes from a filamentous fungal  
 CC cell, discovering new genes, identifying possible functions of unknown  
 CC open reading frames and monitoring gene copy number variation and  
 CC stability. In (M1), one spot on an array equals one gene or open reading  
 CC frame, extensive follow-up characterization is unnecessary since sequence  
 CC information is available and EST microarrays can be organized based on  
 CC function of the gene products. This sequence corresponds to an EST  
 CC sequence of the invention. (Note: this sequence is not given in the  
 CC printed specification but can be obtained in electronic form from the  
 CC USPTO web site seqdata.uspto.gov/sequence.html; Document ID: 6902887B1).  
 XX  
 SQ Sequence 1362 BP; 299 A; 396 C; 348 G; 319 T; 0 U; 0 Other;  
 Query Match 79.0%; Score 15.8; DB 14; Length 1362;  
 Best Local Similarity 89.5%; Pred. No. 6e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 AAACGTGACCTTCACGCCTT 20  
 DB 331 ACACCGACCTTCACGCCTT 349  
 ||| ||||||||||||  
 RESULT 32  
 ACA00487/c  
 ID ACA00487 standard; DNA; 1476 BP.  
 XX  
 AC ACA00487;  
 XX  
 DT 02-JUN-2003 (first entry)  
 XX  
 DE C. glutamicum derived ORF SEQ ID 478.  
 XX  
 KW Coryneform; nucleic acid array; fermentation; culture; ds.  
 XX  
 OS *Corynebacterium glutamicum*.  
 XX  
 PN DE10128510-A1.  
 XX  
 PD 19-DEC-2002.  
 XX  
 PF 13-JUN-2001; 2001DE-01028510.  
 XX  
 PR 13-JUN-2001; 2001DE-01028510.  
 XX  
 XX (DEGS ) DEGUSSA AG.  
 PA  
 PI Farwick M, Moeckel B, Pfefferle W, Bathe B, Huthmacher K;  
 XX WPI; 2003-279970/28.  
 DR  
 XX New nucleic acid array useful for monitoring mRNA expression of  
 PT *Corynebacterium glutamicum* during fermentation, comprising nucleic acid  
 PT

PT from *Corynebacterium glutamicum*.  
XX  
PS Claim 1; Page 220; 709pp; German.  
XX  
CC This invention describes a novel nucleic acid array involving  
CC *Corynebacterium glutamicum* polynucleotides. The arrays are used to  
CC analyse *C. glutamicum*, particularly for monitoring a fermentation process  
CC to determine expression levels of *C. glutamicum* cellular mRNA. Such  
CC monitoring particularly differentiates between expression levels of  
CC different strains of *C. glutamicum* and allows the adjustment of different  
CC culture and fermentation conditions. ACA00010-ACA02188 represent *C.*  
CC *glutamicum* derived polynucleotides described in the disclosure of the  
CC invention  
XX  
SQ Sequence 1476 BP; 265 A; 430 C; 434 G; 347 T; 0 U; 0 Other;  
Query Match 79.0%; Score 15.8; DB 8; Length 1476;  
Best Local Similarity 89.5%; Pred. NO. 6e+02; Length 1476;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 AAACGTGACCTTCACGCCTT 20  
DB 821 ACACCGACCTTCACGCCTT 803  
RESULT 33  
AAH66345/c  
ID AAH66345 standard; DNA; 1491 BP.  
XX  
AC AAH66345;  
XX  
DT 26-SEP-2001 (first entry)  
XX  
DE C *glutamicum* coding sequence fragment SEQ ID NO: 1380.  
XX  
KW *Corynebacterium* bacterium; amino acid synthesis; vitamin; saccharide;  
KW organic acid synthesis; ds.  
XX  
OS *Corynebacterium glutamicum*.  
XX  
FN EP1108790-A2.  
XX  
PD 20-JUN-2001.  
XX  
PF 18-DEC-2000; 2000EP-00127688.  
XX  
PR 16-DEC-1999; 99JP-00377484.  
PR 07-APR-2000; 2000JP-00159162.  
PR 03-AUG-2000; 2000JP-00280988.  
XX  
PA (KYOW ) KYOWA HAKKO KOGYO KK.  
XX  
PI Nakagawa S, Mizoguchi H, Ando S, Hayaishi M, Ochiai K, Yokoi H;  
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
XX  
DR WPI; 2001-376931/40.  
DR P-PSDB; AAG91126.  
XX  
XX Novel polynucleotides derived from *Corynebacterium* bacteria, for identifying  
PT mutation point of a gene, measuring expression of a gene, analyzing  
PT expression profile or pattern of a gene and identifying homologous gene.  
XX  
PS Claim 8; SEQ ID NO 1380; 246pp + Sequence Listing; English.  
XX  
CC The present invention provides a number of nucleotide and protein  
CC sequences from the *Corynebacterium* *Corynebacterium glutamicum*. These  
CC are useful for identifying the mutation point of a gene derived from a  
CC mutant of *Corynebacterium* bacterium, measuring expression amount and analysing  
CC the expression profile or expression pattern of a gene derived from  
CC *Corynebacterium* bacterium, and identifying a homologue of a gene derived from  
CC *Corynebacterium* bacterium. *Corynebacterium* bacteria are useful for producing amino  
CC acids, nucleic acids, vitamins, saccharides and organic acids,  
CC particularly L-lysine. The present sequence is a nucleic acid described

CC in the exemplification of the invention. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from the European Patent Office  
XX  
SQ Sequence 1491 BP; 272 A; 434 C; 436 G; 349 T; 0 U; 0 Other;  
Query Match 79.0%; Score 15.8; DB 5; Length 1491;  
Best Local Similarity 89.5%; Pred. NO. 6.1e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 AAACGTGACCTTCACGCCTT 20  
DB 836 ACACCGACCTTCACGCCTT 818  
RESULT 34  
ADD13197/c  
ID ADD13197 standard; DNA; 1621 BP.  
XX  
AC ADD13197;  
XX  
DT 01-JAN-2004 (first entry)  
XX  
DE C. *glutamicum* stability and folding associated DNA RXA00209.  
XX  
KW ds; gene; genetic stability; DNA repair; recombination; transposition;  
KW gene expression; protein folding; fine chemical production;  
KW lysine production; nucleotide production; nucleoside production;  
KW lipid production; diol production; carbohydrate production;  
KW aromatic compound production; vitamin production; co-factor production;  
KW enzyme production; food; animal feed; cosmetic; pharmaceutical.  
XX  
OS *Corynebacterium glutamicum*.  
XX  
FH Key Location/Qualifiers  
FT CDS 101..1594  
FT /\*tag= a  
XX  
PN WO2003040180-A2.  
XX  
PD 15-MAY-2003.  
XX  
PF 31-OCT-2002; 2002WO-EP012138.  
XX  
PR 05-NOV-2001; 2001DE-01054180.  
XX  
PA (BADI ) BASF AG.  
XX  
PI Zelder O, Pompejus M, Schroeder H, Kroeger B, Kloppeogge C;  
PI Haberhauer G;  
XX  
DR WPI; 2003-505062/47.  
DR P-PSDB; ADD13198.  
XX  
XX New nucleic acid encoding variant forms of proteins required for e.g.  
PT genetic stability and proper protein folding, useful for production of  
PT fine chemicals, specifically lysine, in microorganisms.  
XX  
PS Claim 1; SEQ ID NO 5; 265pp; German.  
XX  
CC This invention describes novel polynucleotides and polypeptides involved  
CC in genetic stability (DNA repair and recombination, transposition of  
CC genetic material), gene expression and folding of proteins in  
CC *Corynebacterium glutamicum*. Polynucleotides are isolated from a nucleic  
CC acid library of *C. glutamicum* then mutated at the specified positions,  
CC cloned and expressed by standard methods. Cells containing vectors that  
CC express the polynucleotides are used for production of fine chemicals,  
CC preferably amino acids and specifically lysine, but more generally  
CC nucleotides, nucleosides, lipids, fatty acids, diols, carbohydrates,  
CC aromatic compounds, vitamins, co-factors and enzymes. These are useful in  
CC the food, animal feed, cosmetics and pharmaceutical industries. The  
CC polynucleotides, optionally as primers and probes, can also be used for  
CC identification and classification of *C. glutamicum* gene manipulation and



CC modulation of metabolic activity. Cells that contain the polynucleotides  
CC of the invention may produce fine chemicals in better yields, with higher  
CC productivity and/or more efficiently.

XX SQ Sequence 1621 BP; 301 A; 463 C; 473 G; 384 T; 0 U; 0 Other;

Query Match 79.0%; Score 15.8; DB 10; Length 1621;  
Best Local Similarity 89.5%; Pred. No. 6.1e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAACGACCTTCACGCCTT 20  
| | | | | | | | | | | | | | | | | |  
DB 936 ACACGACCTTCACGCCTT 918

RESULT 35  
ACL29777  
ID ACL29777 standard; cDNA; 2652 BP.  
AC ACL29777;  
XX 02-JUN-2005 (first entry)  
DT  
XX Rice abiotic stress responsive polynucleotide SEQ ID NO:3733.  
DE  
XX ss; abiotic stress tolerance; transgenic plant; cereal;  
KW agriculture.  
KW  
XX Oryza sativa.  
OS  
XX WO2003008540-A2.  
PN  
XX 30-JAN-2003.  
PD  
XX  
XX 21-JUN-2002; 2002WO-US019668.  
PF  
XX  
XX 22-JUN-2001; 2001US-0300112P.  
PR  
XX 24-AUG-2001; 2001US-0314662P.  
PR  
XX 26-SEP-2001; 2001US-0325277P.  
PR  
XX 21-NOV-2001; 2001US-0332132P.  
PR

XX (SYGN) SYNGENTA PARTICIPATIONS AG.

XX Krepis J, Briggs SP, Cooper B, Glazebrook J, Goff SA, Katagiri F;  
XX Moughamer T, Provart N, Ricke D, Zhu T;  
XX WPI; 2003-248011/24.

XX New stress-responsive nucleic acid, useful for altering the  
XX responsiveness of a plant, e.g. cereal, to an abiotic stress such as cold  
XX stress, salt stress or osmotic stress.

XX Claim 1; SEQ ID NO 3733; 89pp; English.

XX The invention relates to novel abiotic stress responsive polynucleotides  
XX and polypeptides. Also disclosed are vectors, expression cassettes, host  
XX cells, and plants containing such polynucleotides. Also disclosed are  
XX methods for using the polynucleotides and polypeptides to alter the  
XX responsiveness of a plant to abiotic stress. The invention is useful in  
XX agriculture. The nucleic acid is useful for determining whether a test  
XX plant has been exposed to an abiotic stress condition. It is also useful  
XX for selecting an agent that alters abiotic stress regulated  
XX polynucleotide expression in a plant cell, and to identify a homolog or  
XX ortholog to an abiotic stress responsive polynucleotide. The nucleic acid  
XX molecule and the polypeptide encoded by it are useful in altering the  
XX responsiveness of a plant to an abiotic stress, such as cold stress, salt  
XX stress, osmotic stress or any of their combinations. The present sequence  
XX is used in the exemplification of the invention

XX SQ Sequence 2652 BP; 719 A; 543 C; 610 G; 780 T; 0 U; 0 Other;

Query Match 79.0%; Score 15.8; DB 11; Length 2652;  
Best Local Similarity 89.5%; Pred. No. 6.5e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 AAACGACCTTCACGCCTT 20  
| | | | | | | | | | | | | | | | | |  
DB 1718 AAAGTGACCTTCAAGCCTT 1736

RESULT 36  
ABN80098/C  
ID ABN80098 standard; DNA; 5368 BP.

XX AC ABN80098;  
XX  
XX 15-JUL-2002 (first entry)  
DT  
XX Human chemically modified disease associated gene SEQ ID NO 115.

XX Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;  
XX heart disease; epilepsy; histone deacetylation; muscular dystrophy;  
KW dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;  
KW antidiabetic; cytostatic; anticovulsant; ds.  
XX

OS Homo sapiens.  
OS Synthetic.

XX WO200200927-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP007536.

XX 30-JUN-2000; 2000DE-01032529.

XX 01-SEP-2000; 2000DE-01043826.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130908/17.

XX Novel nucleic acid useful for diagnosis and therapy of diseases  
XX associated with development genes such as diabetes, comprises a sequence  
XX of a segment of chemically pretreated DNA of genes associated with  
XX development.

XX Claim 1; SEQ ID NO 115; 27pp; English.

XX The invention relates to a nucleic acid (I) comprising a sequence at  
XX least 18 bases in length of a segment of chemically pretreated DNA (II)  
XX of genes associated with development selected from 87 genes listed in the  
XX specification such as ACCPN, ADFN, or AFD1 and comprising one of 350  
XX sequences (ABN79984-ABN80333) or their complements. The invention is  
XX useful for the diagnosis or therapy of diseases associated with  
XX development genes, in particular disease related to homeobox containing  
XX genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes  
XX associated with congenital heart disease, epilepsy, diseases related to  
XX histone deacetylation, Curran syndrome, diseases related with the  
XX development of the brain and limb girdle muscular dystrophy and dwarfism.  
XX Oligomers specific to each of the genes are useful for detecting the  
XX methylation state of all CpG dinucleotides within the 350 sequences or  
XX (II) and their complementary sequences, as primer oligonucleotides for  
XX the amplification of the 350 sequences, (II) and/or their complements and  
XX as oligomer probes for detecting the cytosine methylation state and/or  
XX single nucleotide polymorphisms (SNPs). Note: The sequence data for this  
XX patent did not form part of the printed specification but is based on  
XX sequence information supplied to Derwent by the European Patent Office

XX SQ Sequence 5368 BP; 1272 A; 208 C; 1745 G; 2143 T; 0 U; 0 Other;

Query Match 79.0%; Score 15.8; DB 6; Length 5368;  
Best Local Similarity 89.5%; Pred. No. 7.2e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
QY      2 AACTGACCTTCAGGCCTT 20
DE      ||||| ||||| ||||| |||||
DB      2021 AACTAACCTTCAGGACTT 2003

RESULT 37
ABL32545/C
ID      ABL32545 standard; DNA; 5882 BP.
XX
AC      ABL32545;
XX
DT      26-MAR-2002 (first entry)
XX
DE      Human immune system associated gene SEQ ID NO: 518.
XX
KW      Human; immune system disease; cytosine methylation; antiasthmatic;
KW      antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW      neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW      antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW      antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW      acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW      neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW      ds.
XX
OS      Homo sapiens.
XX
PN      WO200200928-A2.
XX
PD      03-JAN-2002.
XX
PF      02-JUL-2001; 2001WO-EP007537.
XX
PR      30-JUN-2000; 2000DE-01032529.
XX
PT      01-SEP-2000; 2000DE-01043826.
XX
PA      (EPG-) EPIGENOMICS AG.
XX
PI      Olek A, Piepenbrock C, Berlin K;
XX      WPI; 2002-130909/17.
XX
PT      Nucleic acid comprising fragment of chemically modified gene, useful for
PT      diagnosis and treatment of diseases associated with abnormal cytosine
PT      methylation.
XX
PS      Claim 1; SEQ ID NO 518; 32pp + Sequence Listing; German.
XX
CC      The present invention provides a number of human immune system associated
CC      genes which are modified by the methylation of cytosines. The sequences
CC      can be used in the diagnosis and treatment of immune system disorders,
CC      including eye diseases such as retinopathy, neovascular glaucoma and
CC      macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC      leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC      rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC      diseases. The present sequence is a gene of the invention
XX
SQ      Sequence 5882 BP; 1546 A; 166 C; 1395 G; 2773 T; 0 U; 2 Other;

Query Match      79.0%; Score 15.8; DB 6; Length 5882;
Best Local Similarity 89.5%; Pred. No. 7.3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 AACTGACCTTCAGGCCTT 20
DB      1425 AACCAGCCTTCACACCTT 1407

RESULT 38
ABD33357
ID      ABD33357 standard; DNA; 38538 BP.
XX
AC      ABD33357;
XX

QY      18-NOV-2004 (first entry)
XX
DE      Human cancer-associated (CA) gene HD07-065.
XX
KW      Human; cancer-associated protein; CAP; cancer-associated gene; CA; gene;
KW      ds; cancer; cytostatic.
XX
OS      Homo sapiens.
XX
PN      WO2004058146-A2.
XX
PD      15-JUL-2004.
XX
PF      15-DEC-2003; 2003WO-US040081.
XX
PR      17-DEC-2002; 2002US-00322281.
XX
PA      (SAGR-) SAGRES DISCOVERY INC.
XX
PI      Morris DW, Malandro MS;
XX      WPI; 2004-499109/47.
XX
PT      Novel human cancer associated protein encoded within open reading frame
PT      of cancer associated gene, useful as targets for diagnosing cancer.
XX
PS      Claim 16; SEQ ID NO 448; 182pp; English.
XX
CC      The invention relates to cancer-associated proteins (CAP) and the cancer-
CC      associated (CA) nucleic acids encoding them. The invention also relates
CC      to a method for treating cancers involving administering to a patient an
CC      inhibitor of CAP, and a method of screening for anticancer activity in a
CC      potential drug involving providing a cell that expresses a CA gene,
CC      contacting a tissue sample derived from a cancer cell with an anticancer
CC      drug candidate and monitoring the effect of the anticancer drug candidate
CC      on expression of the CA gene. The CAP proteins are useful for detecting
CC      cancer associated with expression of a CAP protein in a test cell sample
CC      and for screening for a bioactive agent capable of modulating the
CC      activity of a CAP protein. The CA nucleic acids are useful for diagnosing
CC      cancer, involving determining the expression of a CA nucleic acid in a
CC      tissue. This sequence represents a human CA gene of the invention. Note:
CC      The sequence data for this patent did not form part of the printed
CC      specification, but was obtained in electronic format directly from WIPO
CC      at ftp.wipo.int/pub/published_pct_sequences
XX
SQ      Sequence 38538 BP; 9418 A; 9893 C; 9523 G; 9684 T; 0 U; 20 Other;

Query Match      79.0%; Score 15.8; DB 13; Length 38538;
Best Local Similarity 89.5%; Pred. No. 9.3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 GAACTGACCTTCAGGCCT 19
DB      32081 GAACTGACCTTCAGGCCT 32099

RESULT 39
ACF67367_16/c
Continuation (17 of 57) of ACF67367 from base 1600001 (Photorhabdus luminescens nucleoti
WP Sequence split into 57 fragments LOCUS ACF67367 Accession ACF67367
WP Fragment Name      Begin      End
WP ACF67367_00      1      110000
WP ACF67367_01      100001      210000
WP ACF67367_02      200001      310000
WP ACF67367_03      300001      410000
WP ACF67367_04      400001      510000
WP ACF67367_05      500001      610000
WP ACF67367_06      600001      710000
WP ACF67367_07      700001      810000
WP ACF67367_08      800001      910000
WP ACF67367_09      900001     1010000
WP ACF67367_10     1000001     1110000
WP ACF67367_11     1100001     1210000
```

WP ACF67367\_12 1200001 1310000  
WP ACF67367\_13 1300001 1410000  
WP ACF67367\_14 1400001 1510000  
WP ACF67367\_15 1500001 1610000  
WP ACF67367\_16 1600001 1710000  
WP ACF67367\_17 1700001 1810000  
WP ACF67367\_18 1800001 1910000  
WP ACF67367\_19 1900001 2010000  
WP ACF67367\_20 2000001 2110000  
WP ACF67367\_21 2100001 2210000  
WP ACF67367\_22 2200001 2310000  
WP ACF67367\_23 2300001 2410000  
WP ACF67367\_24 2400001 2510000  
WP ACF67367\_25 2500001 2610000  
WP ACF67367\_26 2600001 2710000  
WP ACF67367\_27 2700001 2810000  
WP ACF67367\_28 2800001 2910000  
WP ACF67367\_29 2900001 3010000  
WP ACF67367\_30 3000001 3110000  
WP ACF67367\_31 3100001 3210000  
WP ACF67367\_32 3200001 3310000  
WP ACF67367\_33 3300001 3410000  
WP ACF67367\_34 3400001 3510000  
WP ACF67367\_35 3500001 3610000  
WP ACF67367\_36 3600001 3710000  
WP ACF67367\_37 3700001 3810000  
WP ACF67367\_38 3800001 3910000  
WP ACF67367\_39 3900001 4010000  
WP ACF67367\_40 4000001 4110000  
WP ACF67367\_41 4100001 4210000  
WP ACF67367\_42 4200001 4310000  
WP ACF67367\_43 4300001 4410000  
WP ACF67367\_44 4400001 4510000  
WP ACF67367\_45 4500001 4610000  
WP ACF67367\_46 4600001 4710000  
WP ACF67367\_47 4700001 4810000  
WP ACF67367\_48 4800001 4910000  
WP ACF67367\_49 4900001 5010000  
WP ACF67367\_50 5000001 5110000  
WP ACF67367\_51 5100001 5210000  
WP ACF67367\_52 5200001 5310000  
WP ACF67367\_53 5300001 5410000  
WP ACF67367\_54 5400001 5510000  
WP ACF67367\_55 5500001 5610000  
WP ACF67367\_56 5600001 5648894

Query Match 79.0%; Score 15.8; DB 10; Length 110000;  
Best Local Similarity 89.5%; Pred. No. 1.1e+03;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACGCCT 19  
|||||  
Db 65273 GAAACTGACCTTCACGCCT 65255

RESULT 40  
ADY25743/c  
ID ADY25743 standard; DNA; 209613 BP.  
XX AC ADY25743;  
XX AC  
DT 05-MAY-2005 (first entry)  
XX  
DE Uridine phosphorylase related nucleic acid #7.  
XX  
KW ss; Cytostatic; Gene therapy; Antisense Therapy; UP;  
KW uridine phosphorylase; beta-catenin pathway; modulator; cancer.  
XX  
OS Homo sapiens.  
XX  
PN WO2005017119-A2.  
XX  
PD 24-FEB-2005.

XX 12-AUG-2004; 2004WO-US026339.  
PF  
XX 14-AUG-2003; 2003US-0495172P.  
PR  
XX (EXEL-) EXELIXIS INC.  
PA  
XX Francis-Lang H, Winter CG, Ventura RBA, Heuer TS, Lickteig K;  
PI  
XX WPI; 2005-182359/19.  
DR GENBANK; 4156143.  
XX  
PT Identifying candidate beta catenin pathway modulating agents useful for  
PT diagnosing or treating, e.g. cancer, comprises screening for agents that  
PT modulate the activity of Uridine Phosphorylase.  
XX  
XX Disclosure: SEQ ID NO 7; 181pp; English.  
XX  
XX This sequence represents an UP (uridine phosphorylase) related nucleic  
CC acid. The method of the invention for identifying a candidate beta-  
CC catenin pathway modulating agent comprises screening for agents that  
CC modulate the activity of UP. The method comprises providing an assay  
CC system comprising a UP polypeptide or nucleic acid; contacting the assay  
CC system with a test agent under conditions where, but for the presence of  
CC the test agent, the system provides a reference activity; and detecting a  
CC test agent-biased activity of the assay system, where a difference  
CC between the test agent-biased activity and the reference activity  
CC identifies the test agent as a candidate beta-catenin pathway modulating  
CC agent. The methods of the invention are useful for diagnosing or treating  
CC cancer or for identifying modulators of beta-catenin pathway, which may  
CC be utilized as therapeutic targets for disorders associated with  
CC defective beta-catenin function, such as cancer.  
XX  
SQ Sequence 209613 BP; 64051 A; 43057 C; 41139 G; 61366 T; 0 U; 0 Other;  
  
Query Match 79.0%; Score 15.8; DB 14; Length 209613;  
Best Local Similarity 89.5%; Pred. No. 1.2e+03;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 GAAACTGACCTTCACGCCT 19  
|||||  
Db 37473 GAAACAGACCTTCACGTCT 37455  
  
Search completed: November 20, 2005, 17:43:44  
Job time : 327.938 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2005, 16:52:01 ; Search time 2991.11 Seconds  
(without alignments)  
312.841 Million cell updates/sec

Title: US-10-627-757-21

Perfect score: 20  
Sequence: 1 gaaactgaccttcacgcctt 20

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST.\*  
1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_est3.\*  
4: gb\_est4.\*  
5: gb\_est5.\*  
6: gb\_est6.\*  
7: gb\_est7.\*  
8: gb\_est8.\*  
9: gb\_est9.\*  
10: gb\_est10.\*  
11: gb\_est11.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	17.4	87.0	590	1	AV670322
C 2	17.4	87.0	617	7	CV070718
C 3	17.4	87.0	617	7	CV526995
C 4	17.4	87.0	903	8	CV098206
C 5	17.4	85.0	844	2	BG645018
C 6	16.8	84.0	297	1	BB135169
C 7	16.8	84.0	423	1	AV795955
C 8	16.8	84.0	516	9	BZ164303
C 9	16.8	84.0	576	5	BW176013
C 10	16.8	84.0	589	5	BW130956
C 11	16.8	84.0	603	9	BH275027
C 12	16.8	84.0	611	10	CW781607
C 13	16.8	84.0	619	11	CR874465
C 14	16.8	84.0	620	1	AI981709
C 15	16.8	84.0	628	5	BW344946
C 16	16.8	84.0	635	5	BW035817
C 17	16.8	84.0	637	1	AV859516
C 18	16.8	84.0	659	6	CA665161
C 19	16.8	84.0	674	5	BW006943
C 20	16.8	84.0	674	5	BW460402
C 21	16.8	84.0	717	5	BW374626
C 22	16.8	84.0	723	5	BW066673

23	16.8	84.0	727	5	BW399042
24	16.8	84.0	788	5	BW010505
25	16.8	84.0	793	5	BW163232
26	16.8	84.0	799	5	BW081656
C 27	16.8	84.0	829	10	CW968343
C 28	16.8	84.0	840	10	CW514999
C 29	16.8	84.0	1171	8	DN657911
C 30	16.4	82.0	333	2	BB446157
C 31	16.4	82.0	380	5	BX605863
C 32	16.4	82.0	471	9	AZ710709
C 33	16.4	82.0	558	9	BH793200
C 34	16.4	82.0	647	9	AZ858763
C 35	16.4	82.0	680	10	CL575240
C 36	16.4	82.0	701	10	CL577543
C 37	16.4	82.0	709	9	CC858558
C 38	16.4	82.0	725	9	CE028216
C 39	16.4	82.0	757	10	CL578578
C 40	16.4	82.0	774	5	BUI15246
C 41	16.4	82.0	890	10	DU003601
C 42	16.4	82.0	913	10	AG306604
C 43	16.4	82.0	1003	10	CNS01MQY
C 44	16.4	80.0	843	8	CV908938
C 45	16.4	80.0	923	10	CZ979006

#### ALIGNMENTS

RESULT 1  
AV670322/c

LOCUS AV670322 OLHNI cell line cDNA library (OLb) linear EST 22-SEP-2000  
DEFINITION clone OLb21.09d similar to hypothetical protein F54E7.7  
(Caenorhabditis elegans), mRNA sequence.

ACCESSION AV670322 GI:9936120  
VERSION AV670322.1

KEYWORDS EST.

SOURCE Oryzias latipes (Japanese medaka)

ORGANISM Oryzias latipes

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha; Belontiiformes; Adrianchthyidae; Oryziinae; Oryzias.

REFERENCE 1 (bases 1 to 590)

AUTHORS Naruse,K., Tanaka,M., Shima,A. and Mitani,H.

TITLE Medaka EST Project in University of Tokyo

JOURNAL Unpublished (2000)

COMMENT Contact: Kiyoshi Naruse

Department of Biological Sciences

Graduate School of Science, University of Tokyo

Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan

Tel: 81-3-5841-4443

Fax: 81-3-5841-4410

Email: naruse@biol.s.u-tokyo.ac.jp

This clone was isolated from OLHNI cell line cDNA library (OLb) 5'

end sequences.

FEATURES Location/Qualifiers

source 1..590

/organism="Oryzias latipes"

/mol\_type="mRNA"

/strains="HNI"

/db\_xref="taxon:8090"

/clone="OLb21.09d"

/clone\_lib="OLHNI cell line cDNA library (OLb)"

ORIGIN

Query Match 87.0%; Score 17.4; DB 1; Length 590;

Best Local Similarity 94.7%; Pred. No. 6e+02;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACGCCT 19

|||||

DB 279 GAAACTGACCTTCACGCCT 261

```

RESULT 2
CV070718/c
LOCUS
DEFINITION
CV070718      617 bp      mRNA      linear      EST 24-AUG-2004
CS_gil_30G04_M13Reverse Blue crab gill, normalized Callinectes
sapidus cDNA_clone CS_gil_30G04 5' similar to pir[D96711
hypothetical protein F24J5.8 - imported - Arabidopsis thaliana.
Score = 60.1 bits (144), Expect = 3e-08, mRNA sequence.
CV070718
CV070718.1   GI:515333882
EST.
Callinectes sapidus (blue crab)
ORGANISM
Callinectes sapidus
Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
Eubranchyura; Portunoidae; Portunidae; Callinectes.
1 (bases 1 to 617)
Shaffer,T.H., Coblenz,F.E. and Towle,D.W.
Expressed sequence tags from normalized cDNA libraries prepared
from gill and hypodermis tissues of the blue crab, Callinectes
sapidus
JOURNAL
COMMENT
Unpublished (2004)
Contact: Thomas H. Shafer
Department of Biological Sciences
University of North Carolina Wilmington
601 S. College Rd, Wilmington, NC 28403, USA
Tel: 910-962-7275
Fax: 910-962-4066
Email: shafert@uncw.edu
Plate: 30 row: G column: 04
Seq primer: M13 Reverse
High quality sequence stop: 485.
Location/Qualifiers
1. .617
/organism="Callinectes sapidus"
/mol_type="mRNA"
/db_xref="taxon:6763"
/clone="CS_gil_30G04"
/tissue_type="Pooled anterior and posterior gills from
crabs acclimated to salinities of 35 and 5 parts per
thousand"
/dev_stage="Adult intermolt"
/clone_lib="Blue crab gill, normalized"
/note="Vector: pCMV Sport 6.1; Total RNA samples were
prepared individually from each tissue, checked for
quality, and then pooled for construction and
normalization of a cDNA library by Invitrogen. Plasmids
were isolated and inserts sequenced from their 5'-ends by
the Blue Crab Molecular Genetics Laboratory at the
University of North Carolina Wilmington. Traces were
trimmed, compared (BLASTx) to NCBI non-redundant protein
database as of 19 July 2004, and processed for submission
to dbEST by trace2dbEST software (Parkinson, Anthony and
Blaxter, unpublished software)."
```

```

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Callinectes sapidus (blue crab)
Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
Eubranchyura; Portunoidae; Portunidae; Callinectes.
1 (bases 1 to 617)
Shaffer,T.H., Coblenz,F.E. and Towle,D.W.
Expressed sequence tags from normalized cDNA libraries prepared
from gill and hypodermis tissues of the blue crab, Callinectes
sapidus
JOURNAL
COMMENT
Unpublished (2004)
Contact: Thomas H. Shafer
Department of Biological Sciences
University of North Carolina Wilmington
601 S. College Rd, Wilmington, NC 28403, USA
Tel: 910-962-7275
Fax: 910-962-4066
Email: shafert@uncw.edu
Plate: 07 row: H column: 08
Seq primer: SP6
High quality sequence stop: 511.
Location/Qualifiers
1. .617
/organism="Callinectes sapidus"
/mol_type="mRNA"
/db_xref="taxon:6763"
/clone="CS_GIL_07H08"
/tissue_type="Pooled anterior and posterior gills from
crabs acclimated to salinities of 35 and 5 parts per
thousand"
/dev_stage="Adult intermolt"
/clone_lib="Blue crab gill, normalized"
/note="Vector: pCMV Sport 6.1; Total RNA samples were
prepared individually from each tissue, checked for
quality, and then pooled for construction and
normalization of a cDNA library by Invitrogen. Plasmids
were isolated and inserts sequenced from their 5'-ends by
the Blue Crab Molecular Genetics Laboratory at the
University of North Carolina Wilmington. Traces were
trimmed, compared (BLASTx) to NCBI non-redundant protein
database as of 19 July 2004, and processed for submission
to dbEST by trace2dbEST software (Parkinson, Anthony and
Blaxter, unpublished software)."
```

## ORIGIN

Query Match 87.0%; Score 17.4; DB 7; Length 617;  
 Best Local Similarity 94.7%; Pred. No. 6e+02; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACGCCT 19

||||| |||||||||

Db 378 GAAACTCACCTTCACGCCT 360

## RESULT 4

CV098206

LOCUS

DEFINITION

CV098206

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CV098206  
 EHAH132TR E. histolytica Normalized cDNA library Entamoeba  
 histolytica cDNA, mRNA sequence.  
 CV098206  
 CV098206.1 GI:56626057  
 EST.  
 SOURCE  
 Entamoeba histolytica  
 ORGANISM  
 Entamoeba histolytica  
 Eukaryota; Entamoebidae; Entamoeba.  
 1 (bases 1 to 903)  
 Mann,B., Anderson,I. and Loftus,B.  
 Entamoeba histolytica EST reads  
 Unpublished (2004)  
 Contact: Brendan Loftus

CV526995  
 CS\_GIL\_07H08\_SP6 Blue crab gill, normalized Callinectes sapidus  
 cDNA clone CS\_GIL\_07H08 5' similar to pir[D96711 hypothetical  
 protein F24J5.8 - imported - Arabidopsis thaliana. Score = 58.5  
 bits (140), Expect = 9e-08, mRNA sequence.



/clone\_lib="RIKEN full-length enriched, adult male bone"  
/notes="Site 1: Sali; Site 2: BamHI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN, Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5',  
GAGAGAGAGAGATCAGACATCTTTTCTTTTCTTTTCTT 3'], cDNA was  
prepared by using triphase thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. cDNA went through one round of normalization  
to Rot = 10.0 and subtraction to Rot = 195.0. Second  
strand cDNA was prepared with the primer adapter of  
sequence [5', GAGAGAGAGATCTCGATTAATTAATAATCCCCCCCCCCC  
3']. cDNA was cleaved with XhoI and BamHI. Vector: a  
modified pBluescript KS(+) after bulk excision from Lambda  
PLC-1."

## ORIGIN

```
Query Match      84.0%; Score 16.8; DB 1; Length 297;
Best Local Similarity 90.0%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

Qy 1 GAAACTGACCTTCACGCCTT 20  
||| ||| ||| ||| ||| ||| ||| |||  
D'b 66 GAACCTGACCTTCACACCTT 85

db 66 GAACCTGACCTTCACACCTT 85

## RESULT 7

AV795955/c

LOCUS	AV795955	423 bp	mRNA	linear	EST 29-MAR-2002
DEFINITION	AV795955	RAFL8	Arabidopsis thaliana	cdna clone	RAFL08-19-P05 3', mRNA sequence.

ACCESSION VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE AUTHORS
AV759595		AV759595.1	GI:19829938	
EST.				
			Arabidopsis thaliana (thale cress)	
			Arabidopsis thaliana	
			Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
			Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;	
			rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.	
			1 (bases 1 to 423)	
			Seki, M., Naruseaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M.,	
			Oono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y.,	
			Arakawa, T., Shibata, K., Shingawa, A., Muramatsu, M., Hayashizaki, Y.	
			and Shinozaki, K.	

## FEATURES

source

```
Query Match      84.0%; Score 16.8; DB 9; Length 516;
Best Local Similarity 90.0%; Pred. No. 1.2e+03;
Matches 18: Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

Qy 1 GAAACTGACCTTCACGCCCTT 20  
|||  
Db 159 GAAACTGACCTTCATGCCCTT 170

Db  
159 GAAACTGACTTTCATGCCCTT 178

## RESULT 9

dehydration-treated (1, 2, 5, 10, 24 hr)"

## ORIGIN

Query Match 84.0%; Score 16.8; DB 1; Length 423;  
Best Local Similarity 90.0%; Pred. No. 1.2e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels

Qy 1 GAAACTGACCTTTCAGCCCTT 20  
|||||  
Db 114 GAAATTGAACCTTTCAGCCCTT 95

Db 114 GAAATTGAACCTTCACGCCTT 95

## RESULT 8

BZ164303

LOCUS	BZ164303	516 bp	DNA	linear	GSS 11-OCT-2000
DEFINITION	CH230-388010.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone CH230-388010 genomic survey sequence.				

ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
BZ164303	BZ164303.1	GI:23805321	GSS.	
			Rattus norvegicus (Norway rat)	
			Rattus norvegicus	
			Rattus norvegicus	
			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	

## FEATURES

source

```

FEATURES
source
Class: LNC ends.
Location/Qualifiers
1. .516
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SsNHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-388O10"
/sex="Female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 2"
/notes="Vector: pTARBAC1.3; Site 1: MboI;
CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
Pieter de Jong"

```

Qy 1 GAAACTGACCTTCACGCCCTT 20  
|||  
Db 159 GAAACTGACCTTCATGCCCTT 170

Db  
159 GAAACTGACTTTCATGCCCTT 178

## RESULT 9



```

BW176013
LOCUS      BW176013      576 bp      mRNA      linear      EST 04-NOV-2002
DEFINITION      intestinalis cDNA clone rc1ht010j12 3', mRNA sequence.
ACCESSION      BW176013
VERSION        BW176013.1      GI:24565937
KEYWORDS
SOURCE
ORGANISM      Ciona intestinalis
               Ciona intestinalis
               Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
               Phlebobranchia; Cionidae; Ciona.
REFERENCE
AUTHORS      Satou,Y., Shin-i.T., Kohara,Y. and Satoh,N.
TITLE         Expressed genes in Ciona intestinalis (2002c)
JOURNAL       Unpublished (2002)
COMMENT      Contact: Nori Satoh
               Department of Zoology
               Kyoto University
               Sakyo-Ku, Kyoto, Kyoto 606-8502, Japan
               Tel: 81-75-753-4081
               Fax: 81-75-705-1113
               Email: satoh@ascidian.zool.kyoto-u.ac.jp.
               Location/Qualifiers
               1..576
                   /organism="Ciona intestinalis"
                   /mol_type="mRNA"
                   /db_xref="taxon:7719"
                   /clone="rc1ht010j12"
                   /tissue_type="heart"
                   /clone_lib="Nori Satoh unpublished cDNA library, heart"

FEATURES
source
               Query Match      84.0%;      Score 16.8;      DB 5;      Length 576;
               Best Local Similarity 90.0%;      Pred. No. 1.2e+03;
               Matches 18;      Conservative 0;      Mismatches 2;      Indels 0;      Gaps 0;

ORIGIN
      1      GAAACTGACCTTCACGCCTT 20
      ||||| ||||| ||||| |||||
      488      GAAACGGACCTTCACGCTT 507

      1      GAAACTGACCTTCACGCCTT 20
      ||||| ||||| ||||| |||||
      488      GAAACGGACCTTCACGCTT 507

RESULT 10
LOCUS      BW130956      589 bp      mRNA      linear      EST 01-JUN-2005
DEFINITION      Nori Satoh unpublished cDNA library, gastrula and neurula
               Ciona intestinalis cDNA clone rcign025p06 3', mRNA sequence.
ACCESSION      BW130956
VERSION        BW130956.1      GI:24487355
KEYWORDS
SOURCE
ORGANISM      Ciona intestinalis
               Ciona intestinalis
               Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
               Phlebobranchia; Cionidae; Ciona.
REFERENCE
AUTHORS      Satou,Y., Shin-i.T., Kohara,Y. and Satoh,N.
TITLE         Expressed genes in Ciona intestinalis (2002c)
JOURNAL       Unpublished (2002)
COMMENT      Contact: Nori Satoh
               Department of Zoology
               Kyoto University
               Sakyo-Ku, Kyoto, Kyoto 606-8502, Japan
               Tel: 81-75-753-4081
               Fax: 81-75-705-1113
               Email: satoh@ascidian.zool.kyoto-u.ac.jp.
               Location/Qualifiers
               1..589
                   /organism="Ciona intestinalis"
                   /mol_type="mRNA"
                   /db_xref="taxon:7719"
                   /clone="rcign025p06"
                   /tissue_type="whole animal"
                   /dev_stage="gastrula and neurula"
                   /clone_lib="Nori Satoh unpublished cDNA library, gastrula

FEATURES
source
               Query Match      84.0%;      Score 16.8;      DB 5;      Length 603;
               Best Local Similarity 90.0%;      Pred. No. 1.2e+03;
               Matches 18;      Conservative 0;      Mismatches 2;      Indels 0;      Gaps 0;

ORIGIN
      1      GAAACTGACCTTCACGCCTT 20
      ||||| ||||| ||||| |||||
      502      GAAACGGACCTTCACGCTT 521

      1      GAAACTGACCTTCACGCCTT 20
      ||||| ||||| ||||| |||||
      502      GAAACGGACCTTCACGCTT 521

RESULT 11
LOCUS      BH275027      603 bp      DNA      linear      GSS 30-NOV-2001
DEFINITION      CH230-1C15, TJ CHORI-230 Segment 1 Rattus norvegicus genomic clone
               CH230-1C15, genomic survey sequence.
ACCESSION      BH275027
VERSION        BH275027.1      GI:17187429
KEYWORDS
SOURCE
ORGANISM      Rattus norvegicus (Norway rat)
               Rattus norvegicus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
               Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE
AUTHORS      Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K.,
               Shvartsbeyn,A., Gebregorgis,E., Overton,L., Russell,D., Chen,D.,
               Riggs,F., de Jong,P. and Fraser,C.M.
TITLE         Rat BAC End Sequences from Library CHORI-230 EcoRI segment
JOURNAL       Unpublished (1999)
COMMENT      Other GSSs: CH230-1C15.TV
               Contact: Shaying Zhao
               Department of Eukaryotic Genomics
               The Institute for Genomic Research
               9712 Medical Center Dr., Rockville, MD 20850, USA
               Tel: 301 838 0200
               Fax: 301 838 0208
               Email: szhao@tigr.org
               Clones are derived from the rat BAC library CHORI-230
               (http://www.chori.org/bacpac/rat230.htm). For BAC library
               availability, please contact Pieter de Jong (pdejong@mail.cho.org).
               Clones may be purchased from BACPAC Resources
               (http://www.chori.org/bacpac/or ering information.htm). BAC end
               page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
               Plate: 1 row; C column: 15
               Seq primer: SP6
               Class: BAC ends.
               Location/Qualifiers
               1..603
                   /organism="Rattus norvegicus"
                   /mol_type="genomic DNA"
                   /strain="BN/SSNHsd/MCW"
                   /db_xref="taxon:10116"
                   /clone="CH230-1C15"
                   /sex="Female"
                   /cell_type="Brain"
                   /clone_lib="CHORI-230 Segment 1"
                   /note="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;
                   CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by
                   Pieter de Jong"

FEATURES
source
               Query Match      84.0%;      Score 16.8;      DB 9;      Length 603;
               Best Local Similarity 90.0%;      Pred. No. 1.2e+03;
               Matches 18;      Conservative 0;      Mismatches 2;      Indels 0;      Gaps 0;

ORIGIN
      1      GAAACTGACCTTCACGCCTT 20
      ||||| ||||| ||||| |||||
      577      GAAATGACCTTCACCCCTT 596

      1      GAAACTGACCTTCACGCCTT 20
      ||||| ||||| ||||| |||||
      577      GAAATGACCTTCACCCCTT 596

RESULT 12

```

CW781607/c  
 LOCUS CW781607 611 bp DNA linear GSS 19-NOV-2004  
 DEFINITION OP\_Ba0077H13.r OP\_Ba Oryza punctata genomic clone OP\_Ba0077H13  
 3', genomic survey sequence.

ACCESSION CW781607  
 VERSION CW781607.1 GI:55850911  
 KEYWORDS GSS

SOURCE Oryza punctata  
 ORGANISM Oryza punctata

REFERENCE 1 (bases 1 to 611)  
 AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.

REFERENCE 1 (bases 1 to 611)  
 AUTHORS SanMiguel,P., Westerman,R., Kim,H., Yu,Y., Wisotski,M., Yost,D., Stum,D., Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C., Wing,R. and Jackson,S.A.  
 TITLE OMAP Project - Purdue University  
 JOURNAL Unpublished (2004)  
 COMMENT Contact: Scott A. Jackson  
 Jackson Laboratory  
 Purdue University  
 915 W. State St., West Lafayette, IN 47907, USA  
 Tel: 7654963621  
 Fax: 7654967255  
 Email: sjackson@purdue.edu

Basecalling by phred version 0.020425.c. This sequence was derived from the raw sequence read by clipping with Lucy version 1.19s. Bases 166-776 of the raw sequence (length 995) were retained after clipping.  
 Plate: 0077 row: H column: 13  
 Seq primer: CAC TCA TTA GGC ACC CCA  
 Class: BAC ends.  
 Location/Qualifiers  
 1..611  
 /organism="Oryza punctata"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:4537"  
 /clone="OP\_Ba0077H13"  
 /tissue\_type="young leaves"  
 /lab\_host="DH10B-T1 phage resistant"  
 /clone\_lib="OP\_Ba"  
 /note="Vector: pAGIBAC1; Site\_1: HindIII; Site\_2: HindIII"

FEATURES  
 source

Query Match 84.0%; Score 16.8; DB 10; Length 611;  
 Best Local Similarity 90.0%; Pred. No. 1.2e+03;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 GAACTGACCTTCAGCCTT 20  
 ||||||| |||||||  
 Db 133 GAACTGATGTTACGCCTT 114

ORIGIN

Query Match 84.0%; Score 16.8; DB 10; Length 611;  
 Best Local Similarity 90.0%; Pred. No. 1.2e+03;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAACTGACCTTCAGCCTT 20  
 ||||||| |||||||  
 Db 133 GAACTGATGTTACGCCTT 114  
 RESULT 13  
 LOCUS CR874465 619 bp DNA linear GSS 19-NOV-2004  
 DEFINITION Sus scrofa BES, genomic survey sequence.  
 ACCESSION CR874465  
 VERSION CR874465.1 GI:55872711  
 KEYWORDS GSS; Bac-end sequence BES; Genome Survey Sequence.  
 SOURCE Sus scrofa (pig)

ORGANISM Sus scrofa

REFERENCE 1 (bases 1 to 619)  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae; Sus.

REFERENCE 1 (bases 1 to 619)  
 AUTHORS Rogel-Gaillard,C., Bourgeaux,N., Billault,A., Vaiman,M. and Chardon,P.

TITLE Construction of a swine BAC library: application to the characterization and mapping of porcine type C endoviral elements  
 JOURNAL Cytogenet. Cell Genet. 85 (3-4), 205-211 (1999)  
 PUBMED 10449899

REFERENCE 2 (bases 1 to 619)

AUTHORS Chardon,P., Iannuccelli,N., Roig,A., Dossat,C., Demats,J., Rogel-Gaillard,C., Roy,A., Schibler,L. and Milan,D.  
 TITLE A physical map of the swine genome  
 JOURNAL Unpublished

REFERENCE 3 (bases 1 to 619)  
 AUTHORS Genoscope.

TITLE Direct Submission  
 JOURNAL Submitted (18-NOV-2004)

Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)  
 - Web : www.genoscope.cns.fr

FEATURES  
 source

1..619  
 /organism="Sus scrofa"  
 /mol\_type="genomic DNA"  
 /strain="Large White"  
 /db\_xref="taxon:9823"  
 /clone="b10812F05"  
 /sex="male"  
 /cell\_type="fibroblast"  
 /clone\_lib="SBAB"  
 /note="genoscope sequence ID : IH0AAA79D05RM1"

ORIGIN

Query Match 84.0%; Score 16.8; DB 11; Length 619;  
 Best Local Similarity 90.0%; Pred. No. 1.2e+03;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 GAACTGACCTTCAGCCTT 20  
 ||||||| |||||||  
 Db 356 GAACTGATGTTACGCCTT 375

RESULT 14  
 LOCUS AI981709 620 bp mRNA linear EST. 07-MAY-2001

DEFINITION pat.pk0063.d10.f chicken activated T cell cDNA Gallus gallus cdna clone pat.pk0063.d10.f 5', mRNA sequence.

ACCESSION AI981709  
 VERSION AI981709.1 GI:5884737  
 KEYWORDS EST.

SOURCE Gallus gallus (chicken)  
 ORGANISM Gallus gallus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 620)  
 AUTHORS Tirunagaru,V.G., Sofer,L., Cui,J. and Burnside,J.

TITLE An expressed sequence tag database of T-cell-enriched activated chicken splenocytes: sequence analysis of 5251 clones

JOURNAL Genomics 66 (2), 144-151 (2000)  
 PUBMED 10860659

COMMENT Contact: Joan Burnside  
 Molecular Endocrinology  
 University of Delaware  
 40 Townsend Hall, Newark, DE 19717, USA  
 Tel: 302 831-1345  
 Fax: 302-831-3411  
 Email: joan@udel.edu, www.chickest.udel.edu  
 Seq primer: 17.

FEATURES  
 source

1..620  
 /organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9031"  
 /clone="pat.pk0063.d10.f"  
 /sex="male"  
 /cell\_type="Con A-activated splenic T cell"  
 /lab\_host="E.coli TOP10 F"  
 /clone\_lib="Chicken activated T cell cDNA"  
 /note="Vector: pCDNA3"

ORIGIN

Query Match 84.0%; Score 16.8; DB 1; Length 620;

```

Best Local Similarity 90.0%; Pred. No. 1.2e+03; Mismatches 2; Indels 0; Gaps 0;
Matches 18; Conservative 0;

QY 1 GAAACTGACCTTCACGCCTT 20
    ||||| ||||| ||||| |||||
Db 511 GAAACTGAGCGTCACGCCTT 530

RESULT 15
BW344946/c
LOCUS
DEFINITION
  628 bp mRNA linear EST 27-MAY-2004
  Yutaka Satou unpublished cDNA library, embryo whole animal
  Ciona intestinalis cDNA clone ciem829b20 5', mRNA sequence.
ACCESSION
  BW344946
VERSION
  EST.
KEYWORDS
  Ciona intestinalis
SOURCE
  Ciona intestinalis
  Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
  Phlebobranchia; Cionidae; Ciona.
REFERENCE
  1 (bases 1 to 628)
  Satou,Y., Shin-i,T., Kohara,Y. and Satoh,N.
  Expressed genes in Ciona intestinalis (2004)
  Unpublished (2004)
JOURNAL
  COMMENT
  Contact: Yutaka Satou
  Department of Zoology
  Kyoto University
  Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
  Tel: 81-75-753-4095
  Fax: 81-75-705-1113
  Email: yutaka@ascidian.zool.kyoto-u.ac.jp.

FEATURES
    source
    1..628
    /organism="Ciona intestinalis"
    /mol_type="mRNA"
    /db_xref="taxon:7719"
    /clone="ciem829b20"
    /tissue_type="whole animal"
    /dev_stage="embryo"
    /clone_lib="Yutaka Satou unpublished cDNA library, embryo
    whole animal"

ORIGIN
    Query Match 84.0%; Score 16.8; DB 5; Length 628;
    Best Local Similarity 90.0%; Pred. No. 1.2e+03;
    Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

    QY 1 GAAACTGACCTTCACGCCTT 20
        ||||| ||||| ||||| |||||
    Db 333 GAAACGGACCTTCACGCTT 314

RESULT 16
BW035817/c
LOCUS
DEFINITION
  635 bp mRNA linear EST 13-OCT-2002
  Nori Satoh unpublished cDNA library, blood cells Ciona
  intestinalis cDNA clone cibd029o06 5', mRNA sequence.
ACCESSION
  BW035817
VERSION
  EST.
KEYWORDS
  Ciona intestinalis
SOURCE
  Ciona intestinalis
  Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
  Phlebobranchia; Cionidae; Ciona.
REFERENCE
  1 (bases 1 to 635)
  Satou,Y., Satake,M., Azumi,K., Nonaka,M., Shin-i,T., Kohara,Y. and
  Satoh,N.
  Expressed genes in Ciona intestinalis (2002)
  Unpublished (2002)
JOURNAL
  COMMENT
  Contact: Nori Satoh
  Department of Zoology
  Kyoto University
  Sakyo-ku, Kyoto, Kyoto 606-8502, Japan

FEATURES
    source
    1..637
    /organism="Ciona intestinalis"
    /mol_type="mRNA"
    /db_xref="taxon:7719"
    /clone="rcicl17i03"
    /tissue_type="whole animal"
    /dev_stage="cleaving embryo"
    /clone_lib="Nori Satoh unpublished cDNA library, cleavage
    stage embryo"

ORIGIN
    Query Match 84.0%; Score 16.8; DB 1; Length 637;
    Best Local Similarity 90.0%; Pred. No. 1.2e+03;
    Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

    QY 1 GAAACTGACCTTCACGCCTT 20
        ||||| ||||| ||||| |||||
    Db 482 GAAACGGACCTTCACGCTT 501

RESULT 18
CA665161/c
LOCUS
DEFINITION
  659 bp mRNA linear EST 24-NOV-2002
  wtkl.p0009.b7 wtkl Triticum aestivum cDNA clone wtkl.pk0009.b7 5'
  end, mRNA sequence.
ACCESSION
  CA665161

```

```

Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: sato@ascidian.zool.kyoto-u.ac.jp.

FEATURES
    source
    1..635
    /organism="Ciona intestinalis"
    /mol_type="mRNA"
    /db_xref="taxon:7719"
    /clone="cibd029o06"
    /tissue_type="blood cells"
    /clone_lib="Nori Satoh unpublished cDNA library, blood
    cells"

ORIGIN
    Query Match 84.0%; Score 16.8; DB 5; Length 635;
    Best Local Similarity 90.0%; Pred. No. 1.2e+03;
    Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

    QY 1 GAAACTGACCTTCACGCCTT 20
        ||||| ||||| ||||| |||||
    Db 534 GAAACGGACCTTCACGCTT 515

RESULT 17
AV859516
LOCUS
DEFINITION
  637 bp mRNA linear EST 26-MAY-2005
  Nori Satoh unpublished cDNA library, cleavage stage embryo
  Ciona intestinalis cDNA clone rcicl17i03 3', mRNA sequence.
ACCESSION
  AV859516
VERSION
  EST.
KEYWORDS
  Ciona intestinalis
SOURCE
  Ciona intestinalis
  Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
  Phlebobranchia; Cionidae; Ciona.
REFERENCE
  1 (bases 1 to 637)
  Satoh,N., Satou,Y., Kohara,Y. and Shin-i,T.
  Expressed genes in Ciona intestinalis
  Unpublished (2000)
JOURNAL
  COMMENT
  Contact: Nori Satoh
  Department of Zoology
  Kyoto University
  Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
  Tel: 81-75-753-4081
  Fax: 81-75-705-1113
  Email: sato@ascidian.zool.kyoto-u.ac.jp.

FEATURES
    source
    1..637
    /organism="Ciona intestinalis"
    /mol_type="mRNA"
    /db_xref="taxon:7719"
    /clone="rcicl17i03"
    /tissue_type="whole animal"
    /dev_stage="cleaving embryo"
    /clone_lib="Nori Satoh unpublished cDNA library, cleavage
    stage embryo"

ORIGIN
    Query Match 84.0%; Score 16.8; DB 1; Length 637;
    Best Local Similarity 90.0%; Pred. No. 1.2e+03;
    Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

    QY 1 GAAACTGACCTTCACGCCTT 20
        ||||| ||||| ||||| |||||
    Db 482 GAAACGGACCTTCACGCTT 501

RESULT 18
CA665161/c
LOCUS
DEFINITION
  659 bp mRNA linear EST 24-NOV-2002
  wtkl.p0009.b7 wtkl Triticum aestivum cDNA clone wtkl.pk0009.b7 5'
  end, mRNA sequence.
ACCESSION
  CA665161

```

```

VERSION      CA665161.1  GI:25243686
KEYWORDS     EST.
SOURCE       Triticum aestivum (bread wheat)
ORGANISM     Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
              Poideae; Triticeae; Triticum.
REFERENCE    1 (bases 1 to 659)
AUTHORS      Tingey,S.V., Powell,W., Wolters,P., Dolan,M., Hainey,C., Yuan,Z.,
              Miao,G., Caraher,N. and Hanafey,M.K.
TITLE        DuPont Wheat cDNA Sequence
JOURNAL      Unpublished (2002)
COMMENT      Contact: Scott V. Tingey
              Crop Genetics
              E. I. DuPont de Nemours and Company
              1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
              Tel: 302-631-2602
              Fax: 302-631-2607
              Email: Scott.V.Tingey@USA.dupont.com
              Seq primer: M13.
              Location/Qualifiers
                source
                  1..859
                  /organism="Triticum aestivum"
                  /mol_type="mRNA"
                  /cultivar="Stephens"
                  /db_xref="taxon:4565"
                  /clone="wlki.pk0009.b7"
                  /tissue_type="leaf"
                  /clone_lib="wlki"
                  /note="Vector: pBluescript SK+; Site 1: EcoRI; Site 2:
                  XhoI; Wheat (Triticum aestivum L.) seedlings 1 hr after
                  treatment with
                  6-iodo-3-propyl-2-propyloxy-4(3H)-quinazolinone"

ORIGIN
Query Match      84.0%; Score 16.8; DB 6; Length 659;
Best Local Similarity 90.0%; Pred. No. 1.3e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACGCCTT 20
   |||||||
Db 91 GACACTGACCTTCACGCCTT 72

RESULT 19
BW006943
LOCUS          674 bp mRNA linear EST 12-OCT-2002
DEFINITION    intestinalis cDNA clone rcibd029006 3', mRNA sequence.
ACCESSION     BW006943
VERSION       BW006943.1 GI:23922576
KEYWORDS      EST.
SOURCE        Ciona intestinalis
ORGANISM      Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
              Phlebobranchia; Cionidae; Ciona.
REFERENCE     Satou,Y., Satake,M., Azumi,K., Nonaka,M., Shin-i,T., Kohara,Y. and
              Satoh,N.
TITLE         Expressed genes in Ciona intestinalis (2002)
JOURNAL       Unpublished (2002)
COMMENT       Contact: Nori Satoh
              Department of Zoology
              Kyoto University
              Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
              Tel: 81-75-753-4081
              Fax: 81-75-705-1113
              Email: satoheascidian.zool.kyoto-u.ac.jp.
              Location/Qualifiers
                source
                  1..674
                  /organism="Ciona intestinalis"
                  /mol_type="mRNA"
                  /db_xref="taxon:7719"
                  /clone="cijv034f14"
                  /tissue_type="whole animal"
                  /dev_stage="juvenile"
                  /clone_lib="Nori Satoh unpublished cDNA library, juvenile
                  whole animal"

FEATURES
source
  1..674
  /organism="Ciona intestinalis"
  /mol_type="mRNA"
  /db_xref="taxon:7719"
  /clone="cijv034f14"
  /tissue_type="whole animal"
  /dev_stage="juvenile"
  /clone_lib="Nori Satoh unpublished cDNA library, juvenile
  whole animal"

ORIGIN
Query Match      84.0%; Score 16.8; DB 5; Length 674;
Best Local Similarity 90.0%; Pred. No. 1.3e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACGCCTT 20
   |||||||
Db 452 GAAACGGACCTTCACGCTT 471

RESULT 21
BW374626
LOCUS          717 bp mRNA linear EST 28-MAY-2004
DEFINITION    Yutaka Satou unpublished cDNA library, adult digestive
              gland Ciona intestinalis cDNA clone cig805d09 3', mRNA sequence.
ACCESSION     BW374626
VERSION       BW374626
KEYWORDS      EST.
SOURCE        Ciona intestinalis
ORGANISM      Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
              Phlebobranchia; Cionidae; Ciona.
REFERENCE     Satou,Y., Shin-i,T., Kohara,Y. and Satoh,N.
AUTHORS

```

```

/clone="rcibd029006"
/tissue_type="blood cells"
/clone_lib="Nori Satoh unpublished cDNA library, blood
cells"

```

```

Query Match      84.0%; Score 16.8; DB 5; Length 674;
Best Local Similarity 90.0%; Pred. No. 1.3e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACGCCTT 20
   |||||||
Db 500 GAAACGGACCTTCACGCTT 519

```

```

RESULT 20
BW460402
LOCUS          674 bp mRNA linear EST 10-JUN-2004
DEFINITION    BW460402 Nori Satoh unpublished cDNA library, juvenile whole animal
              Ciona intestinalis cDNA clone cijv034f14 3', mRNA sequence.
ACCESSION     BW460402
VERSION       BW460402.1 GI:48595190
KEYWORDS      EST.
SOURCE        Ciona intestinalis
ORGANISM      Ciona intestinalis
              Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
              Phlebobranchia; Cionidae; Ciona.
REFERENCE     1 (bases 1 to 674)
AUTHORS       Satou,Y., Nakayama,A., Shin-i,T., Kohara,Y. and Satoh,N.
TITLE         Expressed genes in Ciona intestinalis (2004b)
JOURNAL       Unpublished (2004)
COMMENT       Contact: Nori Satoh
              Department of Zoology
              Kyoto University
              Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
              Tel: 81-75-753-4081
              Fax: 81-75-705-1113
              Email: satoheascidian.zool.kyoto-u.ac.jp.
              Location/Qualifiers
                source
                  1..674
                  /organism="Ciona intestinalis"
                  /mol_type="mRNA"
                  /db_xref="taxon:7719"
                  /clone="cijv034f14"
                  /tissue_type="whole animal"
                  /dev_stage="juvenile"
                  /clone_lib="Nori Satoh unpublished cDNA library, juvenile
                  whole animal"

```

```

Query Match      84.0%; Score 16.8; DB 6; Length 659;
Best Local Similarity 90.0%; Pred. No. 1.3e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 GAAACTGACCTTCACGCCTT 20
   |||||||
Db 91 GACACTGACCTTCACGCCTT 72

```

```

RESULT 19
BW006943
LOCUS          674 bp mRNA linear EST 12-OCT-2002
DEFINITION    intestinalis cDNA clone rcibd029006 3', mRNA sequence.
ACCESSION     BW006943
VERSION       BW006943.1 GI:23922576
KEYWORDS      EST.
SOURCE        Ciona intestinalis
ORGANISM      Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
              Phlebobranchia; Cionidae; Ciona.
REFERENCE     Satou,Y., Satake,M., Azumi,K., Nonaka,M., Shin-i,T., Kohara,Y. and
              Satoh,N.
TITLE         Expressed genes in Ciona intestinalis (2002)
JOURNAL       Unpublished (2002)
COMMENT       Contact: Nori Satoh
              Department of Zoology
              Kyoto University
              Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
              Tel: 81-75-753-4081
              Fax: 81-75-705-1113
              Email: satoheascidian.zool.kyoto-u.ac.jp.
              Location/Qualifiers
                source
                  1..674
                  /organism="Ciona intestinalis"
                  /mol_type="mRNA"
                  /db_xref="taxon:7719"

```

```

FEATURES
source
  1..674
  /organism="Ciona intestinalis"
  /mol_type="mRNA"
  /db_xref="taxon:7719"

```

```

FEATURES
source
  1..674
  /organism="Ciona intestinalis"
  /mol_type="mRNA"
  /db_xref="taxon:7719"

```

**TITLE** Expressed genes in Ciona intestinalis (2004)  
**JOURNAL** Unpublished (2004)  
**COMMENT** Contact: Yutaka Satou  
 Department of Zoology  
 Kyoto University  
 Sakyo-ku, Kyoto, Kyoto 606-8502, Japan  
 Tel: 81-75-753-4095  
 Fax: 81-75-705-1113  
 Email: yutaka@ascidian.zool.kyoto-u.ac.jp.

**FEATURES**  
 source  
 1..717  
 Location/Qualifiers  
 /organism="Ciona intestinalis"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:7719"  
 /clone="cldg805d09"  
 /tissue\_type="digestive gland"  
 /dev\_stage="adult"  
 /clone\_lib="Yutaka Satou unpublished cDNA library, adult digestive gland"

**ORIGIN**

Query Match 84.0%; Score 16.8; DB 5; Length 717;  
 Best Local Similarity 90.0%; Pred. No. 1.3e+03;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

**QY** 1 GAAACTGACCTTCACGCCTT 20  
 ||||| ||||| ||||| |||||  
**Db** 453 GAAACGGACCTTCACGCTTT 472

**RESULT 22**

BW066673

**LOCUS**

**DEFINITION**

BW066673 Nori Satoh unpublished cDNA library, cleaving embryo Ciona intestinalis cDNA clone rcic1100d02 3', mRNA sequence.

**ACCESSION**

**VERSION**

**KEYWORDS**

**SOURCE**

**ORGANISM**

**REFERENCE**

**AUTHORS**

**TITLE**

**JOURNAL**

**COMMENT**

Expressed genes in Ciona intestinalis (2002c)  
 Unpublished (2002)  
 Contact: Nori Satoh  
 Department of Zoology  
 Kyoto University  
 Sakyo-ku, Kyoto, Kyoto 606-8502, Japan  
 Tel: 81-75-753-4081  
 Fax: 81-75-705-1113  
 Email: sato@ascidian.zool.kyoto-u.ac.jp.

**FEATURES**

source

1..723  
 Location/Qualifiers  
 /organism="Ciona intestinalis"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:7719"  
 /clone="rcic1100d02"  
 /tissue\_type="whole body"  
 /dev\_stage="cleaving embryo"  
 /clone\_lib="Nori Satoh unpublished cDNA library, cleaving embryo"

**ORIGIN**

Query Match 84.0%; Score 16.8; DB 5; Length 723;  
 Best Local Similarity 90.0%; Pred. No. 1.3e+03;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

**QY** 1 GAAACTGACCTTCACGCCTT 20  
 ||||| ||||| ||||| |||||  
**Db** 500 GAAACGGACCTTCACGCTTT 519

**RESULT 23**

BW399042

**LOCUS**

**DEFINITION**

**ACCESSION**

**VERSION**

**KEYWORDS**

**SOURCE**

**ORGANISM**

**REFERENCE**

**AUTHORS**

**TITLE**

**JOURNAL**

**COMMENT**

Expressed genes in Ciona intestinalis (2004)

Unpublished (2004)

Contact: Yutaka Satou

Department of Zoology

Kyoto University

Sakyo-ku, Kyoto, Kyoto 606-8502, Japan

Tel: 81-75-753-4095

Fax: 81-75-705-1113

Email: yutaka@ascidian.zool.kyoto-u.ac.jp.

Location/Qualifiers

1..727

/organism="Ciona intestinalis"

/mol\_type="mRNA"

/db\_xref="taxon:7719"

/clone="ciem829b20"

/tissue\_type="whole animal"

/dev\_stage="embryo"

/clone\_lib="Yutaka Satou unpublished cDNA library, embryo whole animal"

Query Match 84.0%; Score 16.8; DB 5; Length 727;  
 Best Local Similarity 90.0%; Pred. No. 1.3e+03;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

**QY** 1 GAACTGACCTTCACGCCTT 20  
 ||||| ||||| ||||| |||||  
**Db** 464 GAAACGGACCTTCACGCTTT 483

**ORIGIN**

Query Match 84.0%; Score 16.8; DB 5; Length 727;  
 Best Local Similarity 90.0%; Pred. No. 1.3e+03;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

**QY** 1 GAACTGACCTTCACGCCTT 20  
 ||||| ||||| ||||| |||||  
**Db** 464 GAAACGGACCTTCACGCTTT 483

**RESULT 24**

BW010505

**LOCUS**

**DEFINITION**

**ACCESSION**

**VERSION**

**KEYWORDS**

**SOURCE**

**ORGANISM**

**REFERENCE**

**AUTHORS**

**TITLE**

**JOURNAL**

**COMMENT**

Expressed genes in Ciona intestinalis (2002)

Unpublished (2002)

Contact: Nori Satoh

Department of Zoology

Kyoto University

Sakyo-ku, Kyoto, Kyoto 606-8502, Japan

Tel: 81-75-753-4081

Fax: 81-75-705-1113

Email: sato@ascidian.zool.kyoto-u.ac.jp.

Location/Qualifiers

1..788

/organism="Ciona intestinalis"

/mol\_type="mRNA"

Query Match 84.0%; Score 16.8; DB 5; Length 788;  
 Best Local Similarity 90.0%; Pred. No. 1.3e+03;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

**QY** 1 GAACTGACCTTCACGCCTT 20  
 ||||| ||||| ||||| |||||  
**Db** 500 GAAACGGACCTTCACGCTTT 519

BW399042 727 bp mRNA linear EST 28-MAY-2004  
 BW399042 Yutaka Satou unpublished cDNA library, embryo whole animal  
 Ciona intestinalis cDNA clone ciem829b20 3', mRNA sequence.

BW399042 1 GI:47814870

**EST**

**ORGANISM**

**REFERENCE**

**AUTHORS**

**TITLE**

**JOURNAL**

**COMMENT**

Expressed genes in Ciona intestinalis (2004)

Unpublished (2004)

Contact: Yutaka Satou

Department of Zoology

Kyoto University

Sakyo-ku, Kyoto, Kyoto 606-8502, Japan

Tel: 81-75-753-4095

Fax: 81-75-705-1113

Email: yutaka@ascidian.zool.kyoto-u.ac.jp.

Location/Qualifiers

1..727

/organism="Ciona intestinalis"

/mol\_type="mRNA"

/db\_xref="taxon:7719"

/clone="ciem829b20"

/tissue\_type="whole animal"

/dev\_stage="embryo"

/clone\_lib="Yutaka Satou unpublished cDNA library, embryo whole animal"

```

/db_xref="taxon:7719"
/clone="rcibdd003909"
/tissue_type="blood cells"
/clone_lib="Nori Satoh unpublished cDNA library, blood
cells"

ORIGIN
Query Match      84.0%; Score 16.8; DB 5; Length 788;
Best Local Similarity 90.0%; Pred. No. 1.3e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GAAACTGACCTTCACGCCTT 20
||||| ||||| ||||| |||||
Db 500 GAAACGGACCTTCACGCTT 519

RESULT 25
BW163232
LOCUS      BW163232      793 bp      mRNA      linear      EST 03-NOV-2002
DEFINITION      BW163232 Nori Satoh unpublished cDNA library, gonad Ciona
intestinalis cDNA clone rcigd048c24 3', mRNA sequence.
ACCESSION      BW163232
VERSION      BW163232.1 GI:24520457
KEYWORDS      EST.
SOURCE      Ciona intestinalis
ORGANISM      Ciona intestinalis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
Phlebobranchia; Clonidae; Ciona.
REFERENCE      1 (bases 1 to 793)
AUTHORS      Satou,Y., Shin-i.T., Kohara,Y. and Satoh,N.
TITLE      Expressed genes in Ciona intestinalis (2002c)
JOURNAL      Unpublished (2002)
COMMENT      Contact: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-Ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: sato@ascidian.zool.kyoto-u.ac.jp.
Location/Qualifiers
1..793
/mol_type="mRNA"
/db_xref="taxon:7719"
/clone="rcigd048c24"
/tissue_type="gonad"
/clone_lib="Nori Satoh unpublished cDNA library, gonad"

FEATURES
source
1..793
/organism="Ciona intestinalis"
/mol_type="mRNA"
/db_xref="taxon:7719"
/clone="rcigd048c24"
/tissue_type="gonad"
/clone_lib="Nori Satoh unpublished cDNA library, gonad"

ORIGIN
Query Match      84.0%; Score 16.8; DB 5; Length 793;
Best Local Similarity 90.0%; Pred. No. 1.3e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GAAACTGACCTTCACGCCTT 20
||||| ||||| ||||| |||||
Db 501 GAAACGGACCTTCACGCTT 520

RESULT 26
BW081656
LOCUS      BW081656      799 bp      mRNA      linear      EST 27-MAY-2005
DEFINITION      BW081656 Nori Satoh unpublished cDNA library, egg Ciona
intestinalis cDNA clone rcieg085p03 3', mRNA sequence.
ACCESSION      BW081656
VERSION      BW081656.1 GI:24256933
KEYWORDS      EST.
SOURCE      Ciona intestinalis
ORGANISM      Ciona intestinalis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
Phlebobranchia; Clonidae; Ciona.
REFERENCE      1 (bases 1 to 799)
AUTHORS      Satou,Y., Shin-i.T., Kohara,Y. and Satoh,N.
TITLE      Expressed genes in Ciona intestinalis (2002c)

```

```

Unpublished (2002)
Contact: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: sato@ascidian.zool.kyoto-u.ac.jp.
Location/Qualifiers
1..799
/organism="Ciona intestinalis"
/mol_type="mRNA"
/db_xref="taxon:7719"
/clone="rcieg085p03"
/tissue_type="whole animal"
/dev_stage="egg"
/clone_lib="Nori Satoh unpublished cDNA library, egg"

ORIGIN
Query Match      84.0%; Score 16.8; DB 5; Length 799;
Best Local Similarity 90.0%; Pred. No. 1.3e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GAAACTGACCTTCACGCCTT 20
||||| ||||| ||||| |||||
Db 501 GAAACGGACCTTCACGCTT 520

RESULT 27
CW968343/c
LOCUS      CW968343/c      829 bp      DNA      linear      GSS 21-DEC-2004
DEFINITION      AIAA-aaa05d06.g1 Ancylostoma caninum whole genome shotgun library
(AIAAGSS 001) Ancylostoma caninum genomic, genomic survey sequence.
ACCESSION      CW968343
VERSION      CW968343.1 GI:56769070
KEYWORDS      GSS.
SOURCE      Ancylostoma caninum (dog hookworm)
ORGANISM      Ancylostoma caninum
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
Ancylostomatidae; Ancylostomatidae; Ancylostomatinae; Ancylostoma.
REFERENCE      1 (bases 1 to 829)
AUTHORS      Mitreva,M., McCarter,J.P., Pape,D., Ritter,E., Tsagareishvili,R.,
Ronko,I., Martin,J., Wylie,T., Dante,M., Meyer,R., Messina,D.,
Waterston,R.H., Clifton,S.W. and Wilson,R.
TITLE      Genome Survey sequences from the parasitic nematode Ancylostoma
caninum
JOURNAL      Unpublished (2004)
COMMENT      Contact: Mitreva M
Washington University in St. Louis
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: nematode@watson.wustl.edu
Genomic DNA provided by John Hawdon (mtmjmb@gwumc.edu) DNA
sequenced by Washington University Genome Sequencing Center
Class: shotgun.
Location/Qualifiers
1..829
/organism="Ancylostoma caninum"
/mol_type="genomic DNA"
/strain="Baltimore"
/db_xref="taxon:29170"
/dev_stage="Adult"
/lab_host="GS10"
/clone_lib="Ancylostoma caninum whole genome shotgun
library (AIAAGSS 001)"
/notes="Vector: pOTW13; Site 1: BstXI; Site 2: BstXI;
Ancylostoma caninum genomic DNA was randomly sheared,
end-repaired and size fractionated to enrich for 2-4 kb
fragments. Genomic DNA was provided by John Hawdon
(mtmjmb@gwumc.edu) at George Washington University.
Sequencing by Washington University Genome Sequencing

```

```

ORIGIN
Query Match      84.0%; Score 16.8; DB 10; Length 829;
Best Local Similarity 90.0%; Pred. No. 1.3e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACGCCTT 20
    ||||| ||||| ||||| |||||
DB 66 GAAACTTACCTTCACGCCTT 47

RESULT 28
CW514999/C
LOCUS
DEFINITION      840 bp DNA linear GSS 08-OCT-2004
OP_Ba0021M06.f OP_Ba Oryza punctata genomic clone OP_Ba0021M06
5', genomic survey sequence.
ACCESSION      CW514999
VERSION        CW514999.1 GI:53989221
KEYWORDS       GSS.
SOURCE         Oryza punctata
ORGANISM       Oryza punctata
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE      1 (bases 1 to 840)
AUTHORS        SanMiguel,P., Westerman,R., Kim,H., Yu,Y., Wissotski,M., Yost,D.,
                Stum,D., Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C.,
                Hatfield,J., Soderlund,C., Wing,R. and Jackson,S.A.
TITLE          OMAP Project - Purdue University
JOURNAL        Unpublished (2004)
COMMENT        Contact: Scott A. Jackson
                Jackson Laboratory
                915 W. State St., West Lafayette, IN 47907, USA
                Tel: 7654963621
                Fax: 7654967255
                Email: sjackson@purdue.edu
                Basecalling by phred version 0.020425.c. This sequence was derived
                from the raw sequence read by clipping with lucy version 1.19s.
                Bases 38-877 of the raw sequence (length 1375) were retained after
                clipping.
PCR Primers    FORWARD: TAA TAC GAC TCA CTA TAG GG
                BACKWARD: CAC TCA TTA GGC ACC CCA
                Insert Length: 161000 Std Error: 0.00
                Plate: 0021 row: M column: 06
                Seq primer: TAA TAC GAC TCA CTA TAG GG
                Class: BAC ends.
FEATURES       Location/Qualifiers
                1..840
                /organism="Oryza punctata"
                /mol_type="genomic DNA"
                /db_xref="taxon:4537"
                /clone="OP_Ba0021M06"
                /tissue_type="young leaves"
                /lab_host="DH10B-T1 phage resistant"
                /clone_lib="Op_Ba"
                /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"

ORIGIN
Query Match      84.0%; Score 16.8; DB 10; Length 840;
Best Local Similarity 90.0%; Pred. No. 1.3e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACGCCTT 20
    ||||| ||||| ||||| |||||
DB 167 GAAACTGATGTTACGCCTT 148

RESULT 29
DN657911
LOCUS
DEFINITION      1171 bp mRNA linear EST 28-MAR-2005
DN657911
5', mRNA sequence.
ACCESSION      DN657911
VERSION        DN657911.1 GI:61963160
KEYWORDS       EST.
SOURCE         Gasterosteus aculeatus (three spined stickleback)
ORGANISM       Gasterosteus aculeatus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;
Gasterosteidae; Gasterosteus.
REFERENCE      1 (bases 1 to 1171)
AUTHORS        Kingsley,D.M., Peichel,C., Balabhadra,S., Grimwood,J., Dickson,M.,
                Schmutz,J. and Myers,R.M.
TITLE          Expressed sequence tags from Gasterosteus aculeatus
JOURNAL        Unpublished (2003)
COMMENT        Contact: Grimwood, Jane
                Stanford Human Genome Center
                Stanford University School of Medicine
                975 S California Ave, Palo Alto, CA 94304, USA
                Tel: 650 320 5917
                Fax: 650 320 5801
                Email: jane@hgsc.stanford.edu
                Plate: 31
                High quality sequence start: 16
                High quality sequence stop: 813.
                High quality sequence Location/Qualifiers
                1..1171
                /organism="Gasterosteus aculeatus"
                /mol_type="mRNA"
                /strain="Conner Creek sticklebacks, WA USA"
                /db_xref="taxon:69293"
                /clone="CEC31-E12"
                /sex="mixed male and female"
                /tissue_type="skin"
                /dev_stage="adult"
                /lab_host="DH10B (T1 phage resistant)"
                /clone_lib="SHGC-CEC"
                /note="Vector: Express 1; Total and poly A+ RNA was
                isolated from the indicated stickleback tissue, and a cDNA
                library was constructed in the Express 1 plasmid vector by
                Open Biosystems. First strand cDNA synthesis was primed
                with an 54 bp linker primer containing an oligo dT sequence
                preceded by a synthetic NotI site (first strand primer:
                5'-GACTAGTTCTAGATCGGCGGCC(T)25-3'). Following
                second strand synthesis, cDNAs were made blunt at the end
                corresponding to the original 5 prime end of mRNA, and
                cloned directionally into the NotI and EcoRV sites of
                Express 1. Note that the EcoRV site is typically destroyed
                in the blunt end cloning, leaving a junction of the form
                'xxxATC' (where is ATC is the second half of the EcoRV
                site, and xxx is derived from the cDNA sequence). A map of
                the Express 1 vector is available at:
                http://www.openbiosystems.com/cdna_library_construction_fa
                q.php#8 The primary library was transformed and amplified
                in DH10B (T1 phage resistant) bacteria. Clones available
                from Open Biosystems:
                http://www.openbiosystems.com/stickleback"

FEATURES       Location/Qualifiers
                1..1171
                /organism="Gasterosteus aculeatus"
                /mol_type="mRNA"
                /strain="Conner Creek sticklebacks, WA USA"
                /db_xref="taxon:69293"
                /clone="CEC31-E12"
                /sex="mixed male and female"
                /tissue_type="skin"
                /dev_stage="adult"
                /lab_host="DH10B (T1 phage resistant)"
                /clone_lib="SHGC-CEC"
                /note="Vector: Express 1; Total and poly A+ RNA was
                isolated from the indicated stickleback tissue, and a cDNA
                library was constructed in the Express 1 plasmid vector by
                Open Biosystems. First strand cDNA synthesis was primed
                with an 54 bp linker primer containing an oligo dT sequence
                preceded by a synthetic NotI site (first strand primer:
                5'-GACTAGTTCTAGATCGGCGGCC(T)25-3'). Following
                second strand synthesis, cDNAs were made blunt at the end
                corresponding to the original 5 prime end of mRNA, and
                cloned directionally into the NotI and EcoRV sites of
                Express 1. Note that the EcoRV site is typically destroyed
                in the blunt end cloning, leaving a junction of the form
                'xxxATC' (where is ATC is the second half of the EcoRV
                site, and xxx is derived from the cDNA sequence). A map of
                the Express 1 vector is available at:
                http://www.openbiosystems.com/cdna_library_construction_fa
                q.php#8 The primary library was transformed and amplified
                in DH10B (T1 phage resistant) bacteria. Clones available
                from Open Biosystems:
                http://www.openbiosystems.com/stickleback"

ORIGIN
Query Match      84.0%; Score 16.8; DB 8; Length 1171;
Best Local Similarity 90.0%; Pred. No. 1.4e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACGCCTT 20
    ||||| ||||| ||||| |||||
DB 49 GAAACTGACCTTACACCTT 68

RESULT 30
BB446157
LOCUS
DEFINITION      333 bp mRNA linear EST 19-JUL-2000
BB446157
5', mRNA sequence.
ACCESSION      BB446157
VERSION        BB446157.1 GI:1171
KEYWORDS       EST.
SOURCE         RIKEN full-length enriched, 9 days embryo Mus musculus
ORGANISM       RIKEN full-length enriched, 9 days embryo Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;
Gasterosteidae; Gasterosteus.
REFERENCE      1 (bases 1 to 1171)
AUTHORS        Kingsley,D.M., Peichel,C., Balabhadra,S., Grimwood,J., Dickson,M.,
                Schmutz,J. and Myers,R.M.
TITLE          Expressed sequence tags from Gasterosteus aculeatus
JOURNAL        Unpublished (2003)
COMMENT        Contact: Grimwood, Jane
                Stanford Human Genome Center
                Stanford University School of Medicine
                975 S California Ave, Palo Alto, CA 94304, USA
                Tel: 650 320 5917
                Fax: 650 320 5801
                Email: jane@hgsc.stanford.edu
                Plate: 31
                High quality sequence start: 16
                High quality sequence stop: 813.
                High quality sequence Location/Qualifiers
                1..1171
                /organism="Gasterosteus aculeatus"
                /mol_type="mRNA"
                /strain="Conner Creek sticklebacks, WA USA"
                /db_xref="taxon:69293"
                /clone="CEC31-E12"
                /sex="mixed male and female"
                /tissue_type="skin"
                /dev_stage="adult"
                /lab_host="DH10B (T1 phage resistant)"
                /clone_lib="SHGC-CEC"
                /note="Vector: Express 1; Total and poly A+ RNA was
                isolated from the indicated stickleback tissue, and a cDNA
                library was constructed in the Express 1 plasmid vector by
                Open Biosystems. First strand cDNA synthesis was primed
                with an 54 bp linker primer containing an oligo dT sequence
                preceded by a synthetic NotI site (first strand primer:
                5'-GACTAGTTCTAGATCGGCGGCC(T)25-3'). Following
                second strand synthesis, cDNAs were made blunt at the end
                corresponding to the original 5 prime end of mRNA, and
                cloned directionally into the NotI and EcoRV sites of
                Express 1. Note that the EcoRV site is typically destroyed
                in the blunt end cloning, leaving a junction of the form
                'xxxATC' (where is ATC is the second half of the EcoRV
                site, and xxx is derived from the cDNA sequence). A map of
                the Express 1 vector is available at:
                http://www.openbiosystems.com/cdna_library_construction_fa
                q.php#8 The primary library was transformed and amplified
                in DH10B (T1 phage resistant) bacteria. Clones available
                from Open Biosystems:
                http://www.openbiosystems.com/stickleback"

```



cdna clone D030054021 3' similar to AF056574 Mus musculus  
pyrroline-5-carboxylate synthetase short isoform mRNA, mRNA  
sequence.  
BB446157  
BB446157.1 GI:9288769  
EST.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 333)  
Konno.H., Aizawa.K., Akahira.S., Akiyama.J., Arakawa.T.,  
Carninci.P., Endo.T., Fukuda.S., Fukunishi.Y., Hara.A., Hayatsu.N.,  
Hirozane.T., Hori.F., Ishii.Y., Ishikawa.J., Ishikawa.T., Itoh.M.,  
Izawa.M., Kadota.K., Kagawa.I., Kai.C., Kawai.J., Kikuchi.N.,  
Kiyosawa.H., Kojima.Y., Kondo.S., Koya.S., Kurihara.C.,  
Kusakabe.M., Matsuyama.T., Miki.R., Mizuno.Y., Nakamura.M., Oda.H.,  
Okazaki.Y., Ono.T., Owa.C., Saito.H., Sakai.C., Sato.K.,  
Shibata.K., Shibata.Y., Shigemoto.Y., Shingawa.A., Shiraki.T.,  
Sogabe.Y., Sugahara.Y., Suzuki.H., Suzuki.H., Tagawa.A.,  
Takahashi.F., Tomimaga.N., Toya.T., Tsunoda.Y., Watahiki.A.,  
Watanabe.S., Yamamura.T., Yamanaka.I., Yano.R., Yasunishi.A.,  
Yokota.T., Yoshida.K., Yoshiki.A., Yoshino.M., Muramatsu.M. and  
Hayashizaki.Y.  
RIKEN mouse ESTs (Konno.H., et al.)  
Unpublished (2000)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-resgsc.riken.jp, URL: http://genome.gsc.riken.jp/  
Carninci.P., Nishiyama.Y., Westover.A., Itoh.M., Nagaoka.S.,  
Sasaki.N., Okazaki.Y., Muramatsu.M. and Hayashizaki.Y.  
Thermostabilization and thermoactivation of thermolabile enzymes by  
trehalose and its application for the synthesis of full length  
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Itoh.M., Kiteunai.T., Akiyama.J., Shibata.K., Izawa.M., Kawai.J.,  
Tomaru.Y., Carninci.P., Shibata.Y., Ozawa.Y., Muramatsu.M.,  
Okazaki.Y. and Hayashizaki.Y.  
Automated filtration-based high-throughput plasmid preparation  
system. Genome Res. 9 (5), 463-470 (1999)  
Carninci.P. and Hayashizaki.Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,  
19-44 (1999)  
Please visit our web site (http://genome.rtc.riken.go.jp) for  
further details.  
Location/Qualifiers  
1..333  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="D030054021"  
/dev\_stage="9 days embryo"  
/lab\_host="DH10B"  
/clone\_lib="RIKEN full-length enriched, 9 days embryo"  
/note="Site 1: Sali; Site 2: BamHI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5'  
GAGAGAGCGCGCGCAACTCGAGTTTTTTTTTTTNN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. Second strand cDNA was prepared with the  
primer adapter of sequence [5'  
GAGAGAGATTCTCGAGTTAAATTAATTCCTCCCTCCCCCCCC 3']. cDNA  
was cleaved with BamHI and XhoI. Vector: a modified

ORIGIN  
Query Match 82.0%; Score 16.4; DB 2; Length 333;  
Best Local Similarity 94.4%; Pred. NO. 1.8e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
pBluescript KS(+) after bulk excision from Lambda PLC I."  
QY 1 GAAACTGACCTTCAGGCC 18  
DB 135 GAATCTGACCTTCAGGCC 152  
RESULT 31  
BX605863/c  
LOCUS BX605863 380 bp mRNA linear EST 07-AUG-2003  
DEFINITION BX605863 Normalized Anopheles Fat Body (NAFB) Library Anopheles  
gambiae cDNA clone AGCBH70TR, mRNA sequence.  
ACCESSION BX605863  
VERSION BX605863.1 GI:33495750  
KEYWORDS EST.  
ORGANISM Anopheles gambiae (African malaria mosquito)  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;  
Culicidae; Anophelinae; Anopheles.  
REFERENCE 1 (bases 1 to 380)  
Lobo,N.L., Gardner,M., Romans,P. and Collins,F.H.  
Anopheles gambiae EST, Center for Tropical Disease Research and  
Training  
Unpublished (2003)  
Contact: Frank H. Collins  
Center for Tropical Disease Research and Training  
University of Notre Dame  
Notre Dame, IN 46556, USA  
Tel: 574-631-9245  
Fax: 574-631-3996  
Email: frank.h.collins.75@nd.edu.  
FEATURES  
source  
1..380  
/organism="Anopheles gambiae"  
/mol\_type="mRNA"  
/db\_xref="taxon:7165"  
/clone="AGCBH70TR"  
/lab\_host="E. coli DH10B"  
/clone\_lib="Normalized Anopheles Fat Body (NAFB) Library"  
/note="Vector: p7713B-Pac (Pharmacia) with a modified  
polylinker; Site 1: EcoRI (5'end); Site 2: NotI (3'end); a  
directionally cloned and normalized, oligo-T primed cDNA  
library constructed from equal numbers of 24 and 40 hr  
post Plasmodium inui infection, 4arr and L35 strain adult  
female mosquito abdomens (omitting ovaries and midguts):  
Bonaldo, Lennon & Soares (1996): Normalization and  
Subtraction: Two Approaches To Facilitate Gene Discovery,  
Genome Research 6, 791-806. ESTs sequenced from the M13  
reverse priming site reading from the 5' ends of the cDNAs  
are indicated by 'R' in the clone name."  
ORIGIN  
Query Match 82.0%; Score 16.4; DB 5; Length 380;  
Best Local Similarity 94.4%; Pred. NO. 1.9e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 3 AACTGACCTTCAGGCC 20  
DB 166 AACTGACCTTCAGGCC 149  
RESULT 32  
AZ710709  
LOCUS AZ710709 471 bp DNA linear GSS 24-JAN-2001  
DEFINITION RPCI-24-116P20.TJ RPCI-24 Mus musculus genomic clone  
RPCI-24-116P20, genomic survey sequence.  
ACCESSION AZ710709

```

VERSION      AZ710709.1  GI:12442550
KEYWORDS     Mus musculus (house mouse)
SOURCE       Mus musculus
ORGANISM     Mus musculus
REFERENCE    Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
              Sciurognathi; Muroidea; Muridae; Murinae; Mus.
              1 (bases 1 to 471)
TITLE        Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
JOURNAL       Tsengaye,G., Geer,K., Kroi,M., Shvartsbeyn,A., Gebregeorgis,E.,
COMMENT      Russell,D., de Jong,P. and Fraser,C.M.
              Mouse BAC End Sequences from Library RPCI-24
              Unpublished (1999)
              Other_GSSs: RPCI-24-116P20.TV
              Contact: Shaying Zhao
              Department of Eukaryotic Genomics
              The Institute for Genomic Research
              9712 Medical Center Dr., Rockville, MD 20850, USA
              Tel: 301 838 0200
              Fax: 301 838 0208
              Email: szhao@tigr.org
              Clones are derived from the mouse BAC library RPCI-24. For BAC
              library availability, please contact Pieter de Jong
              (pdejong@mail.cho.org). Clones may be purchased from BACPAC
              Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
              page: http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html
              Plate: 116 row: P column: 20
              Seq primer: SP6
              Class: BAC ends.
FEATURES     Location/Qualifiers
              1..471
                /organism="Mus musculus"
                /mol_type="genomic DNA"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="RPCI-24-116P20"
                /sex="Male"
                /cell_type="Spleen/Brain"
                /clone_lib="RPCI-24"
                /notes="Vector: pTARBAC1; Site 1: BamH1; Site 2: BamH1;
                RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                library was cloned in the pTARBAC1 cloning vector at the
                BamH1 sites using MboI partially digested male C57BL/6J
                DNA."
ORIGIN
Query Match      82.0%; Score 16.4; DB 9; Length 471;
Best Local Similarity 94.4%; Pred. No. 1.9e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3  AACTGACCTTCACGCCTT 20
        |||||
        142 AACTGACCTTCACGCTT 159
        |||||
        142 AACTGACCTTCACGCTT 159

RESULT 33
LOCUS      BH793200
DEFINITION BH0276 Sub-clone library of Mycoplasma haemofelis BAC clones
            Mycoplasma haemofelis genomic, genomic survey sequence.
ACCESSION  BH793200
VERSION     BH793200.1  GI:29374488
KEYWORDS   GSS.
SOURCE     Mycoplasma haemofelis
ORGANISM   Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
REFERENCE  1 (bases 1 to 558)
AUTHORS    Berent,L.M. and Messick,J.B.
TITLE      Physical Map and Genome Sequencing Survey of Mycoplasma haemofelis
            (Haemobartonella felis)
JOURNAL    Infect. Immun. 71 (6), 3657-3662 (2003)
PUBMED     12761157
COMMENT    Contact: Messick, J.B.

```

```

Haemotropic Mycoplasma Genome Project
University of Illinois - Department of Veterinary Pathobiology
2001 S. Lincoln Ave, Urbana, IL 61802, USA
Tel: 2173332008
Fax: 2172447421
Email: jmessick@cvm.uiuc.edu
Plate: BAC 6.88 3002H05
Class: plasmid ends.
FEATURES   Location/Qualifiers
            1..558
              /organism="Mycoplasma haemofelis"
              /mol_type="genomic DNA"
              /strain="Ohio"
              /db_xref="taxon:29501"
              /lab_host="Felis catus"
              /clone_lib="Sub-clone library of Mycoplasma haemofelis BAC
              clones"
              /note="Vector: pGEM-3zf(+); Site 1: HindIII; pBelOBAC11
              clones of M. haemofelis DNA were sub-cloned into
              pGEM-3zf(+) for sequencing."
ORIGIN
Query Match      82.0%; Score 16.4; DB 9; Length 558;
Best Local Similarity 94.4%; Pred. No. 2e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3  AACTGACCTTCACGCCTT 20
        |||||
        524 AACTGCCCTTCACGCCTT 541
        |||||

RESULT 34
LOCUS      AZ858763/c
DEFINITION 2M0164003F Mouse 10kb plasmid UUGCLM library Mus musculus genomic
            clone UUGC2M0164003 F, genomic survey sequence.
ACCESSION  AZ858763
VERSION     AZ858763.1  GI:13052203
KEYWORDS   GSS.
SOURCE     Mus musculus (house mouse)
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
            1 (bases 1 to 647)
            Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
            Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von
            Niederhausern,A. and Wright,D.,Weiss,R.
            Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
            Unpublished (2000)
            Contact: Robert B. Weiss
            University of Utah Genome Center
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: dunn@genetics.utah.edu
            Insert Length: 10000 Std Error: 0.00
            Plate: 0164 row: O column: 03
            Seq primer: CGTTGTAACGACGCCAGT
            Class: plasmid ends
            High quality sequence stop: 647.
FEATURES     Location/Qualifiers
            1..647
              /organism="Mus musculus"
              /mol_type="genomic DNA"
              /strain="C57BL/6J"
              /db_xref="taxon:10090"
              /clone="UUGC2M0164003"
              /sex="Male"

```

/lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
 /clone\_lib="Mouse 10kb plasmid UUGc1m library"  
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnates/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

## ORIGIN

Query Match 82.0%; Score 16.4; DB 9; Length 647;  
 Best Local Similarity 94.4%; Pred. No. 2e+03;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAACTGACCTTCACGCC 18  
 ||| ||||| ||||| |||||  
 Db 261 GAATCTGACCTTCACGCC 244

## RESULT 35

CL575240

LOCUS

DEFINITION OB\_Ba0023G02.r OB\_Ba Oryza brachyantha genomic clone GSS 10-JUN-2004  
 OB\_Ba0023G02 3', genomic survey sequence.

ACCESSION CL575240

VERSION CL575240.1 GI:48532752

KEYWORDS GSS.

SOURCE Oryza brachyantha

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzeae; Oryza.

1 (bases 1 to 680)

REFERENCE Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,  
 Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.

OMAP Project

Unpublished (2004)

Contact: Rod A. Wing

Arizona Genomics Institute

University of Arizona

Forbes Building Room 303, Tucson, AZ 85721-0036, USA

Tel: 520 626 9595

Fax: 520 621 1259

Email: rwing@genome.arizona.edu

PCR Primers

FORWARD: TAA TAC GAC TCA CTA TAG GG

BACKWARD: CAC TCA TTA GGC ACC CCA

Insert Length: 145 Std Error: 0.00

Plate: 0023 row: G column: 02

Seq primer: CAC TCA TTA GGC ACC CCA

Class: BAC ends.

Location/Qualifiers

1..680

/organism="Oryza brachyantha"

/mol\_type="genomic DNA"

/db\_xref="taxon:4533"

/clone="OB\_Ba0023G02"

/tissue\_type="leaves"

/dev\_stage="mature"

/lab\_host="DH10B"

/clone\_lib="OB\_Ba"  
 /note="Vector: PAGIBAC1; Site\_1: HindIII; Site\_2: HindIII"

## ORIGIN

Query Match 82.0%; Score 16.4; DB 10; Length 680;  
 Best Local Similarity 94.4%; Pred. No. 2e+03;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAACGTGACCTTCACGCC 19

||||| ||||| ||||| |||||

Db 195 AAACGTGACCTTCACACCT 212

## RESULT 36

CL577543

LOCUS

DEFINITION OB\_Ba0031N08.r OB\_Ba Oryza brachyantha genomic clone GSS 17-JUN-2005  
 OB\_Ba0031N08 3', genomic survey sequence.

ACCESSION CL577543

VERSION CL577543.2 GI:67947270

KEYWORDS GSS.

SOURCE Oryza brachyantha

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzeae; Oryza.

1 (bases 1 to 701)

REFERENCE Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,  
 Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.

OMAP Project

Unpublished (2004)

On Jun 17, 2005 this sequence version replaced gi:48535055.

Contact: Rod A. Wing

Arizona Genomics Institute

University of Arizona

Forbes Building Room 303, Tucson, AZ 85721-0036, USA

Tel: 520 626 9595

Fax: 520 621 1259

Email: rwing@genome.arizona.edu

PCR Primers

FORWARD: TAA TAC GAC TCA CTA TAG GG

BACKWARD: CAC TCA TTA GGC ACC CCA

Insert Length: 145 Std Error: 0.00

Plate: 0031 row: N column: 08

Seq primer: CAC TCA TTA GGC ACC CCA

Class: BAC ends.

Location/Qualifiers

1..701

/organism="Oryza brachyantha"

/mol\_type="genomic DNA"

/db\_xref="taxon:4533"

/clone="OB\_Ba0031N08"

/tissue\_type="leaves"

/dev\_stage="mature"

/lab\_host="DH10B"

/clone\_lib="OB\_Ba"

/note="Vector: PAGIBAC1; Site\_1: HindIII; Site\_2: HindIII"

## ORIGIN

Query Match 82.0%; Score 16.4; DB 10; Length 701;  
 Best Local Similarity 94.4%; Pred. No. 2.1e+03;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAACGTGACCTTCACGCC 19

||||| ||||| ||||| |||||

Db 195 AAACGTGACCTTCACACCT 212

## RESULT 37

CC858558/c

LOCUS

DEFINITION ND.L.62C4.SP6 Notre Dame Liverpool Aedes aegypti genomic clone GSS 24-JUL-2003  
 CC858558 709 bp DNA linear  
 ND.L.62C4.SP6 Notre Dame Liverpool-62C4, genomic survey sequence.

```

ACCESSION      CC858558
VERSION        CC858558.1  GI:33217535
KEYWORDS       GSS.
SOURCE         Aedes aegypti (yellow fever mosquito)
ORGANISM       Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
               Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
               Culicidae; Culicinae; Culicini; Aedes; Stegomyia.
REFERENCE      1 (bases 1 to 709)
AUTHORS        Loftus,B., Shetty,J., Knudson,D. and Severson,D.
TITLE          BAC end sequencing of Aedes aegypti
JOURNAL        Unpublished (2003)
COMMENT        Other_GSSs: NDL.62C4.T7
               Contact: Brendan Loftus
               Department of Eukaryotic Genomics
               TIGR
               9712 Medical Center Drive, Rockville, MD 20850, USA
               Tel: 301-838-3543
               Fax: 301-838-0208
               Email: enta@tigr.org
               Library was provided by David Severson
               Seq primer: SP6
               Class: BAC ends.

FEATURES       Location/Qualifiers
               source
               1..709
               /organism="Aedes aegypti"
               /mol_type="genomic DNA"
               /strain="Liverpool"
               /db_xref="taxon:7159"
               /clone="Notre Dame Liverpool-62C4"
               /clone_lib="Notre Dame Liverpool"
               /note="Vector: pECBAC1; Site 1: Hind III; The library was
               prepared from whole body tissue of newly hatched L1 larvae
               by David Severson at the University of Notre Dame and
               Hongbin Zhang"

ORIGIN
Query Match      82.0%; Score 16.4; DB 9; Length 709;
Best Local Similarity 94.4%; Pred. NO. 2.1e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2  AAATGACCTTCAGCCT 19
Db      426 AAATGACCTTCAGCCT 409

RESULT 38
CE028216
LOCUS
DEFINITION      tigr-gss-dog-1700037512808 Dog Library Canis familiaris genomic,
genomic survey sequence.
ACCESSION      CE028216
VERSION        CE028216.1  GI:35047720
KEYWORDS       GSS..
SOURCE         Canis familiaris (dog)
ORGANISM       Canis familiaris
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
               Canis.
REFERENCE      1 (bases 1 to 725)
AUTHORS        Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
               Rusch,D.B., Delcher,A.L., Pop,M., Wang,M., Fraser,C.M. and
               Venter,J.C.
TITLE          The dog genome: survey sequencing and comparative analysis
JOURNAL        Science 301 (5641), 1898-1903 (2003)
PUBLISHED      14512627
COMMENT        Contact: Kirkness EF
               The Institute for Genomic Research
               Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
               Rockville, MD 20850, USA
               Tel: 301-838-0200
               Fax: 301-838-0208
               Email: ekirknes@tigr.org

```

```

Class: shotgun.
Location/Qualifiers
source
1..725
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site 1: BstXI; Libraries were prepared from
peripheral blood"

ORIGIN
Query Match      82.0%; Score 16.4; DB 9; Length 725;
Best Local Similarity 94.4%; Pred. NO. 2.1e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2  AAATGACCTTCAGCCT 19
Db      645 AAATGACCTTCAGCCT 662

RESULT 39
CE028216
LOCUS
DEFINITION      OB_Ba0034009.r OB_Ba Oryza brachyantha genomic clone
OB_Ba0034009 3', genomic survey sequence.
ACCESSION      CE028216
VERSION        CE028216.1  GI:48536090
KEYWORDS       GSS.
SOURCE         Oryza brachyantha
               Oryza brachyantha
               Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
               Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
               Ehrhartoideae; Oryzeae; Oryza.
REFERENCE      1 (bases 1 to 757)
AUTHORS        Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
               Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
TITLE          OMAP Project
JOURNAL        Unpublished (2004)
COMMENT        Contact: Rod A. Wing
               Arizona Genomics Institute
               University of Arizona
               Forbes Building Room 303, Tucson, AZ 85721-0036, USA
               Tel: 520 626 9595
               Fax: 520 621 1259
               Email: rwing@genome.arizona.edu
               PCR Primers
               FORWARD: TAA TAC GAC TCA CTA TAG GG
               BACKWARD: CAC TCA TTA GGC ACC CCA
               Insert Length: 145000 Std Error: 0.00
               Plate: 0034 row: 0 column: 09
               Seq primer: CAC TCA TTA GGC ACC CCA
               Class: BAC ends.
Location/Qualifiers
source
1..757
/organism="Oryza brachyantha"
/mol_type="genomic DNA"
/db_xref="taxon:4533"
/clone="OB_Ba0034009"
/tissue_type="leaves"
/dev_stage="mature"
/lab_host="DH10B"
/clone_lib="OB_Ba"
/note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"

ORIGIN
Query Match      82.0%; Score 16.4; DB 10; Length 757;
Best Local Similarity 94.4%; Pred. NO. 2.1e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2  AAATGACCTTCAGCCT 19
Db      195 AAATGACCTTCAGCCT 212

```

RESULT 40  
BU315246/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
PUBMED  
COMMENT  
FEATURES  
Source

BU315246 774 bp mRNA linear EST 28-NOV-2002  
603852773F1 CSEQCHN62 Gallus gallus cDNA clone ChEST851b19 5', mRNA  
sequence.  
BU315246  
BU315246.1 GI:25823247  
EST.  
Gallus gallus (chicken)  
Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
1 (bases 1 to 774)  
Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,  
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.  
A Comprehensive Collection of Chicken cDNAs  
Curr. Biol. 12 (22), 1965-1969 (2002)  
12445392  
Contact: Simon Hubbard  
Department of Biomolecular Sciences  
University of Manchester Institute of Science and Technology  
(UMIST)  
PO Box 88, Manchester, M60 1QD, UK  
Tel: 01612008930  
Fax: 01612360409  
Email: Simon.Hubbard@umist.ac.uk.  
Location/Qualifiers  
1..774  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/strain="White Leghorn, Hisex"  
/db\_xref="taxon:9031"  
/clone="ChEST851b19"  
/dev\_stage="36"  
/lab\_host="DH10B"  
/clone\_lib="CSEQCHN62"  
/note="Organ: heads; Vector: pBluescript II KS(+); Site\_1:  
EcoRI; Site\_2: NotI; This normalized library was  
constructed from 1 million independent clones. cDNA  
synthesis was initiated using an oligo(dT) primer, using  
methylated C in the first strand synthesis reaction.  
Following this first strand reaction, double-stranded cDNA  
was bluntended, ligated to NotI adapters, digested with  
EcoRI, size-selected, and cloned into the NotI and EcoRI  
compatible sites of a custom modified MCS of the  
pBluescript (KS+) vector. The library was normalized in 2  
rounds using conditions adapted from Soares et al., PNAS  
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6  
(1996): 791, except that a significantly longer  
reannealing hybridization was used."

ORIGIN  
Query Match 82.0%; Score 16.4; DB 5; Length 774;  
Best Local Similarity 94.4%; Pred. No. 2.1e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 3 AACTGACCTTCACGCCTT 20  
|||||  
Db 614 AACTGACCTTCACGCCTT 597

Search completed: November 20, 2005, 21:55:29  
Job time : 2995.11 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2005, 17:14:50 ; Search time 91.1111 Seconds  
(without alignments)  
390.196 Million cell updates/sec

Title: US-10-627-757-21  
Perfect score: 20  
Sequence: 1 gaactgaccttcacgcctt 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_NA:\*

1: /cgn2\_6/prodata/1/ina/1 COMB.seq:\*

2: /cgn2\_6/prodata/1/ina/5 COMB.seq:\*

3: /cgn2\_6/prodata/1/ina/6A COMB.seq:\*

4: /cgn2\_6/prodata/1/ina/6B COMB.seq:\*

5: /cgn2\_6/prodata/1/ina/H COMB.seq:\*

6: /cgn2\_6/prodata/1/ina/PCITUS COMB.seq:\*

7: /cgn2\_6/prodata/1/ina/PP COMB.seq:\*

8: /cgn2\_6/prodata/1/ina/RE COMB.seq:\*

9: /cgn2\_6/prodata/1/ina/baCkfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	15.8	79.0	510	3	US-09-573-080A-370
2	15.8	79.0	641	3	US-09-533-559-595
3	15.8	79.0	1362	3	US-09-533-559-5889
4	15.8	79.0	1956	3	US-09-614-221A-248
5	15.8	79.0	24735	3	US-09-949-016-12850
6	15.4	77.0	2553	3	US-09-489-039A-6319
7	15.2	76.0	228	3	US-09-252-991A-5904
8	15.2	76.0	336	3	US-09-821-167-5
9	15.2	76.0	348	3	US-09-513-999C-14818
10	15.2	76.0	593	3	US-09-533-559-255
11	15.2	76.0	601	3	US-09-949-016-43843
12	15.2	76.0	601	3	US-09-949-016-43883
13	15.2	76.0	601	3	US-09-949-016-52460
14	15.2	76.0	601	3	US-09-949-016-52505
15	15.2	76.0	601	3	US-09-949-016-67365
16	15.2	76.0	601	3	US-09-949-016-67366
17	15.2	76.0	601	3	US-09-949-016-131787
18	15.2	76.0	601	3	US-09-949-016-131787
19	15.2	76.0	879	3	US-09-949-016-182863
20	15.2	76.0	1254	3	US-09-328-352-432
21	15.2	76.0	1409	3	US-09-252-991A-5889
22	15.2	76.0	1409	3	US-09-640-173-193
23	15.2	76.0	1409	3	US-09-713-550-193
24	15.2	76.0	1409	3	US-09-825-294-193
	15.2	76.0	1409	3	US-09-970-966-193

c 25	15.2	76.0	1636	3	US-09-023-655-1005	Sequence 1005, Ap
c 26	15.2	76.0	2075	3	US-10-104-047-1872	Sequence 1872, Ap
c 27	15.2	76.0	2565	3	US-10-104-047-1814	Sequence 1814, Ap
c 28	15.2	76.0	3469	3	US-10-104-047-413	Sequence 413, App
c 29	15.2	76.0	5838	2	US-08-578-096A-1	Sequence 1, Appli
c 30	15.2	76.0	5838	3	US-09-240-426-1	Sequence 1, Appli
c 31	15.2	76.0	11188	3	US-09-821-167-1	Sequence 1, Appli
c 32	15.2	76.0	11752	3	US-09-949-016-11756	Sequence 11756, A
c 33	15.2	76.0	11865	3	US-09-949-016-15297	Sequence 15297, A
c 34	15.2	76.0	11865	3	US-09-949-016-15298	Sequence 15298, A
c 35	15.2	76.0	23383	3	US-09-949-016-13726	Sequence 13726, A
c 36	15.2	76.0	25905	3	US-09-949-016-16788	Sequence 16788, A
c 37	15.2	76.0	49164	3	US-09-949-016-12985	Sequence 12985, A
c 38	15.2	76.0	49164	3	US-09-949-016-12986	Sequence 12986, A
c 39	15.2	76.0	49164	3	US-09-949-016-13262	Sequence 13262, A
c 40	15.2	76.0	49164	3	US-09-949-016-13263	Sequence 13263, A
c 41	15.2	76.0	57267	3	US-09-949-016-11899	Sequence 11899, A
c 42	15.2	76.0	101015	3	US-09-949-016-16981	Sequence 16981, A
c 43	15.2	76.0	147321	3	US-09-949-016-15450	Sequence 15450, A
c 44	15.2	76.0	156942	3	US-09-949-016-12227	Sequence 12227, A
c 45	15.2	76.0	156950	3	US-09-949-016-15946	Sequence 15946, A

ALIGNMENTS

RESULT 1

US-09-573-080A-370  
; Sequence 370, Application US/09573080A  
; Patent No. 6828097  
; GENERAL INFORMATION:  
; APPLICANT: JOAN, KNOLL  
; APPLICANT: ROGAN, PETER  
; TITLE OF INVENTION: SINGLE COPY GENOMIC HYBRIDIZATION PROBES AND METHOD OF GENERATING  
; FILE REFERENCE: 30307  
; CURRENT APPLICATION NUMBER: US/09/573,080A  
; CURRENT FILING DATE: 2000-05-16  
; NUMBER OF SEQ ID NOS: 479  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 370  
; LENGTH: 510  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: repeat region  
; LOCATION: (1)..(510)  
; OTHER INFORMATION: ltr10b  
; PUBLICATION INFORMATION:  
; AUTHORS: Jurka, J; Malchewicz, J; Milosavljevic, A  
; TITLE: Prototypic sequences for human repetitive DNA  
; JOURNAL: Journal of Molecular Evolution  
; VOLUME: 35  
; ISSUE: 4  
; DATE: 1992-10-  
; PAGES: 286-291  
; DATABASE ACCESSION NUMBER: Database of repetitive elements (repbase)  
; DATABASE ENTRY DATE: 1996-01-26  
; DATABASE ENTRY DATE: 1996-01-26  
US-09-573-080A-370

Query Match 79.0%; Score 15.8; DB 3; Length 510;  
Best Local Similarity 89.5%; Pred. No. 67;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAACGTGACCTTCACGCCTT 20  
||||| ||||| |||||  
Db 217 AAACGTGACCTTCACGCCTT 235

RESULT 2

US-09-533-559-595  
; Sequence 595, Application US/09533559

```
; Patent No. 6902887
; GENERAL INFORMATION:
; APPLICANT: Randy M. Berka
; APPLICANT: Michael W. Rey
; APPLICANT: Jeffrey R. Shuster
; APPLICANT: Sakari Kauppinen
; APPLICANT: Ib Groth Clausen
; APPLICANT: Peter Bjørke Olsen
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 5849,200-US
; CURRENT APPLICATION NUMBER: US/09/533,559
; CURRENT FILING DATE: 2000-03-22
; EARLIER APPLICATION NUMBER: 09/273,623
; EARLIER FILING DATE: 1999-03-22
; NUMBER OF SEQ ID NOS: 7860
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 595
; LENGTH: 641
; TYPE: DNA
; ORGANISM: Fusarium venenatum
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(641)
; OTHER INFORMATION: n = A,T,C or G
US-09-533-559-595
```

```
Query Match          79.0%; Score 15.8; DB 3; Length 641;
Best Local Similarity 89.5%; Pred. No. 71;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 GAAACTGACCTTCACGCCT 19
Db 469 GAACTGTCCTTCACGACT 487
|||||
```

```
RESULT 3
US-09-533-559-5689
; Sequence 5689, Application US/09533559
; Patent No. 6902887
; GENERAL INFORMATION:
; APPLICANT: Randy M. Berka
; APPLICANT: Michael W. Rey
; APPLICANT: Jeffrey R. Shuster
; APPLICANT: Sakari Kauppinen
; APPLICANT: Ib Groth Clausen
; APPLICANT: Peter Bjørke Olsen
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 5849,200-US
; CURRENT APPLICATION NUMBER: US/09/533,559
; CURRENT FILING DATE: 2000-03-22
; EARLIER APPLICATION NUMBER: 09/273,623
; EARLIER FILING DATE: 1999-03-22
; NUMBER OF SEQ ID NOS: 7860
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5689
; LENGTH: 1362
; TYPE: DNA
; ORGANISM: Aspergillus oryzae
US-09-533-559-5689
```

```
Query Match          79.0%; Score 15.8; DB 3; Length 1362;
Best Local Similarity 89.5%; Pred. No. 82;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 2 AAACGTGACCTTCACGCCTT 20
Db 331 ACACCGACCTTCACGCCTT 349
|||||
```

```
RESULT 4
US-09-614-221A-248
```

```
; Sequence 248, Application US/09614221A
; Patent No. 6723837
; GENERAL INFORMATION:
; APPLICANT: Karunanandaa, Balasulojini
; APPLICANT: Yu, Jaehyuk
; APPLICANT: Kishore, Ganesh M.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED
; FILE REFERENCE: 16516,075
; CURRENT APPLICATION NUMBER: US/09/614,221A
; CURRENT FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/142,981
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 626
; SEQ ID NO 248
; LENGTH: 1956
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-614-221A-248
```

```
Query Match          79.0%; Score 15.8; DB 3; Length 1956;
Best Local Similarity 89.5%; Pred. No. 88;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 GAAACTGACCTTCACGCCT 19
Db 870 GAACTGACTTAACGCCT 888
|||||
```

```
RESULT 5
US-09-949-016-12850
; Sequence 12850, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12850
; LENGTH: 24735
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12850
```

```
Query Match          79.0%; Score 15.8; DB 3; Length 24735;
Best Local Similarity 89.5%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 2 AAACGTGACCTTCACGCCTT 20
Db 10139 AAACGTGATCTTCCGCCTT 10157
|||||
```

```
RESULT 6
US-09-489-039A-6319/c
; Sequence 6319, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709,2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
```



\* Sequence 14818, Application US/09513999C  
; Patent No. 6783961  
; GENERAL INFORMATION:

APPLICANT: Duclert, A.  
APPLICANT: Giordano, J. Y.  
TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.

```

: FILE REFERENCE: 33-0021-180
:
: CURRENT APPLICATION NUMBER: US/09/513,999C
:
: CURRENT FILING DATE: 2000-02-24
:
: PRIOR APPLICATION NUMBER: US 60/122,487
:
: PRIOR FILING DATE: 1999-02-26
:
: NUMBER OF SEQ ID NOS: 36681
:

```

```

, SOFTWARE: Patent.pm
, SEQ ID NO 14818
, LENGTH: 348

```

```

; I1221: DNA
; ORGANISM: Homo sapiens
US-09-513-999C-14818

Query Match          76.0%; Score 15.2; DB 3; Length 348;
Best Local Similarity 85.0%; Pred. NO. 1.3e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

Qy	1	GAACTGACCTTCACGCTT	20
Db	261	GAAATTGACCTTGACGCTT	280

US-09-533-559-255/c  
; Sequence 255, Application US/09533559  
; Patent No. 6902887

GENERAL INFORMATION:  
 APPLICANT: Randy M. Berka  
 APPLICANT: Michael W. Rey  
 APPLICANT: Jeffrey R. Shuster

```

? APPLICANT: Peter Bjarke Olsen
? TITLE OF INVENTION: Methods for Monitoring Multiple Gene
? TITLE OF INVENTION: Expression
? FILE REFERENCE: 5849.200-US
? CURRENT APPLICATION NUMBER: US/09/533,559
? CURRENT FILING DATE: 2000-03-22
? EARLIER APPLICATION NUMBER: 09/273,623
? EARLIER FILING DATE: 1999-03-22
? NUMBER OF SEQ ID NOS: 7860
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 255
? LENGTH: 593
? TYPE: DNA
? ORGANISM: Fusarium venenatum
? FEATURE:
? NAME/KEY: misc feature
? LOCATION: (1)...(593)
? OTHER INFORMATION: n = A,T,C or G
US-09-533-559-255

Query Match          76.0%   Score 15.2   DB 3   Length 593:

```

Matches	17;	Conservative	0;	Mismatches	3;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	----	--------	----	------	----

QY 1 GAAACTGACCTTCACGCCTT 20  
|||||  
Db 438 GAAAGTGACCTTGACGCCT 419

RESULT 11  
US-09-949-016-43843  
; Sequence 43843, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 43843  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-43843

Query Match 76.0%; Score 15.2; DB 3; Length 601;  
Best Local Similarity 85.0%; Pred. No. 1.5e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACGCCTT 20  
|||||  
Db 36 GAAAGTGACCTTCCTGCCTT 55

RESULT 12  
US-09-949-016-43888  
; Sequence 43888, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 43888  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-43888

Query Match 76.0%; Score 15.2; DB 3; Length 601;  
Best Local Similarity 85.0%; Pred. No. 1.5e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACGCCTT 20  
|||||  
Db 36 GAAAGTGACCTTCCTGCCTT 55

RESULT 13  
US-09-949-016-52460  
; Sequence 52460, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 52460  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-52460

Query Match 76.0%; Score 15.2; DB 3; Length 601;  
Best Local Similarity 85.0%; Pred. No. 1.5e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACGCCTT 20  
|||||  
Db 36 GAAAGTGACCTTCCTGCCTT 55

RESULT 14  
US-09-949-016-52505  
; Sequence 52505, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 52505  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-52505

Query Match 76.0%; Score 15.2; DB 3; Length 601;  
Best Local Similarity 85.0%; Pred. No. 1.5e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACGCCTT 20  
|||||  
Db 36 GAAAGTGACCTTCCTGCCTT 55

RESULT 15  
US-09-949-016-67365  
; Sequence 67365, Application US/09949016  
; Patent No. 6812339

```
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 67365
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-67365

Query Match          76.0%; Score 15.2; DB 3; Length 601;
Best Local Similarity 85.0%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACGCCTT 20
   ||||| ||||| ||||| |||||
Db 252 GACACTGACATTCACCTCTT 271

RESULT 16
US-09-949-016-67366
; Sequence 67366, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 67366
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-67366

Query Match          76.0%; Score 15.2; DB 3; Length 601;
Best Local Similarity 85.0%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACGCCTT 20
   ||||| ||||| ||||| |||||
Db 73 GACACTGACATTCACCTCTT 92

RESULT 17
US-09-949-016-131787
; Sequence 131787, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
```

```
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 131787
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-131787

Query Match          76.0%; Score 15.2; DB 3; Length 601;
Best Local Similarity 85.0%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACGCCTT 20
   ||||| ||||| ||||| |||||
Db 350 GAAAAGGACCTTCAGGCCTT 369

RESULT 18
US-09-949-016-182863/c
; Sequence 182863, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 182863
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-182863

Query Match          76.0%; Score 15.2; DB 3; Length 601;
Best Local Similarity 85.0%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACGCCTT 20
   ||||| ||||| ||||| |||||
Db 146 GAAACTGACCTCCCAAGTCTT 127

RESULT 19
US-09-328-352-432/c
; Sequence 432, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 432
; LENGTH: 879
```

```
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-432

Query Match          76.0%; Score 15.2; DB 3; Length 879;
Best Local Similarity 85.0%; Pred. No. 1.6e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCAGCCCTT 20
   ||||| ||||| ||||| |||||
Db 802 GAAGCTGTCTTCAGCCCT 783

RESULT 20
US-09-252-991A-5889/c
; Sequence 5889, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5889
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5889

Query Match          76.0%; Score 15.2; DB 3; Length 1254;
Best Local Similarity 85.0%; Pred. No. 1.7e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCAGCCCTT 20
   ||||| ||||| ||||| |||||
Db 40 GAAGCTGACCTTCAGCCCTT 21

RESULT 21
US-09-640-173-193
; Sequence 193, Application US/09640173
; Patent No. 6613515
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: OVARIAN TUMOR SEQUENCES AND
; TITLE OF INVENTION: METHODS OF USE THEREFOR
; FILE REFERENCE: 210121.484C2
; CURRENT APPLICATION NUMBER: US/09/640,173
; CURRENT FILING DATE: 2000-08-15
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 193
; LENGTH: 1409
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-640-173-193

Query Match          76.0%; Score 15.2; DB 3; Length 1409;
Best Local Similarity 85.0%; Pred. No. 1.8e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCAGCCCTT 20
   ||||| ||||| ||||| |||||
Db 582 GAAATTGACCTTCAGCCCTT 601

RESULT 22
US-09-713-550-193
; Sequence 193, Application US/09713550
; Patent No. 6617109
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.484C4
; CURRENT APPLICATION NUMBER: US/09/713,550
; CURRENT FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 205
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 193
; LENGTH: 1409
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-713-550-193

Query Match          76.0%; Score 15.2; DB 3; Length 1409;
Best Local Similarity 85.0%; Pred. No. 1.8e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCAGCCCTT 20
   ||||| ||||| ||||| |||||
Db 582 GAAATTGACCTTCAGCCCTT 601

RESULT 23
US-09-825-294-193
; Sequence 193, Application US/09825294
; Patent No. 6710170
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Stolk, John A.
; APPLICANT: Algate, Paul A.
; APPLICANT: Fling, Steven P.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.484C5
; CURRENT APPLICATION NUMBER: US/09/825,294
; CURRENT FILING DATE: 2001-04-03
; NUMBER OF SEQ ID NOS: 215
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 193
; LENGTH: 1409
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-825-294-193

Query Match          76.0%; Score 15.2; DB 3; Length 1409;
Best Local Similarity 85.0%; Pred. No. 1.8e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCAGCCCTT 20
   ||||| ||||| ||||| |||||
Db 582 GAAATTGACCTTCAGCCCTT 601

RESULT 24
US-09-970-966-193
; Sequence 193, Application US/09970966
; Patent No. 6720146
; GENERAL INFORMATION:
; APPLICANT: Stolk, John A.
; APPLICANT: Molesh, David Alan
; APPLICANT: Fling, Steven P.
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.484C6
; CURRENT APPLICATION NUMBER: US/09/970,966
```

```
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-432

Query Match          76.0%; Score 15.2; DB 3; Length 879;
Best Local Similarity 85.0%; Pred. No. 1.6e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCAGCCCTT 20
   ||||| ||||| ||||| |||||
Db 802 GAAGCTGTCTTCAGCCCT 783

RESULT 20
US-09-252-991A-5889/c
; Sequence 5889, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5889
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5889

Query Match          76.0%; Score 15.2; DB 3; Length 1254;
Best Local Similarity 85.0%; Pred. No. 1.7e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCAGCCCTT 20
   ||||| ||||| ||||| |||||
Db 40 GAAGCTGACCTTCAGCCCTT 21

RESULT 21
US-09-640-173-193
; Sequence 193, Application US/09640173
; Patent No. 6613515
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: OVARIAN TUMOR SEQUENCES AND
; TITLE OF INVENTION: METHODS OF USE THEREFOR
; FILE REFERENCE: 210121.484C2
; CURRENT APPLICATION NUMBER: US/09/640,173
; CURRENT FILING DATE: 2000-08-15
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 193
; LENGTH: 1409
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-640-173-193

Query Match          76.0%; Score 15.2; DB 3; Length 1409;
Best Local Similarity 85.0%; Pred. No. 1.8e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCAGCCCTT 20
   ||||| ||||| ||||| |||||
Db 582 GAAATTGACCTTCAGCCCTT 601

RESULT 22
US-09-713-550-193
; Sequence 193, Application US/09713550
; Patent No. 6617109
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.484C4
; CURRENT APPLICATION NUMBER: US/09/713,550
; CURRENT FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 205
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 193
; LENGTH: 1409
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-713-550-193

Query Match          76.0%; Score 15.2; DB 3; Length 1409;
Best Local Similarity 85.0%; Pred. No. 1.8e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCAGCCCTT 20
   ||||| ||||| ||||| |||||
Db 582 GAAATTGACCTTCAGCCCTT 601

RESULT 23
US-09-825-294-193
; Sequence 193, Application US/09825294
; Patent No. 6710170
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Stolk, John A.
; APPLICANT: Algate, Paul A.
; APPLICANT: Fling, Steven P.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.484C5
; CURRENT APPLICATION NUMBER: US/09/825,294
; CURRENT FILING DATE: 2001-04-03
; NUMBER OF SEQ ID NOS: 215
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 193
; LENGTH: 1409
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-825-294-193

Query Match          76.0%; Score 15.2; DB 3; Length 1409;
Best Local Similarity 85.0%; Pred. No. 1.8e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCAGCCCTT 20
   ||||| ||||| ||||| |||||
Db 582 GAAATTGACCTTCAGCCCTT 601

RESULT 24
US-09-970-966-193
; Sequence 193, Application US/09970966
; Patent No. 6720146
; GENERAL INFORMATION:
; APPLICANT: Stolk, John A.
; APPLICANT: Molesh, David Alan
; APPLICANT: Fling, Steven P.
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.484C6
; CURRENT APPLICATION NUMBER: US/09/970,966
```

```

: CURRENT FILING DATE: 2001-10-02
: NUMBER OF SEQ ID NOS: 215
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 193
: LENGTH: 1409
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-970-966-193

```

Query Match 76.0%; Score 15.2; DB 3; Length 1409;  
Best Local Similarity 85.0%; Pred. NO. 1.8e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GAAACTGACCTTCACGCCTT 20  
Db 582 GAAATTGACCTTCACGCCTT 601

RESULT 25  
US-09-023-655-1005/c  
; Sequence 1005, Application US/09023655  
; Patent No. 6607879

```
Query Match          76.0%; Score 15.2; DB 3; Length 1636;
Best Local Similarity 85.0%; Pred. No. 1.8e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

Qy 1 GAACTGACCTTCAGCCTT 20

Db 1371 GTATCTGACCTTCAGGCCTT 1352

```

RESULT 26
US-10-104-047-1872/c
; Sequence 1872, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. 6943241el full length cDNA
; FILE REFERENCE: HI-A0105
; CURRENT APPLICATION NUMBER: US/104,047
; CURRENT FILING DATE: 2002-03-25

```

Query Match	Score	DB 3
Best Local Similarity	76.0%	Length 2075;
Matches	85.0%	Pred. No. 1.9e+02;
17; Conservative	0;	Mismatches 3; Indels 0; Gaps 0;

QY 1 GAACTGACCTTCACGCCTT 20

```

RESULT 27
US-10-104-047-1814/c
; Sequence 1814, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. 6943241el full length cDNA
; FILE REFERENCE: HI-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIORITY APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1814
; LENGTH: 2565
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-1814

```

Query Match	76.0%	Score 15.2;	DB 3;	Length 2565;
Best Local Similarity	85.0%	Pred. No. 2e+02;		
Matches 17;	Conservative	0;	Mismatches 3;	Indels 0;
Gaps	0;			

**Qy**            1 GAAACTGACCTTCAGGCCTT 20  
               | | | | | | | | | |  
**pB**            2366 GTATCTGACCTTCAGGCCTT 2347

```

RESULT 28
US-10-104-047-413/c
; Sequence 413, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. 6943241el full length cDNA
; FILE REFERENCE: HA-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
;

```

```
/ NUMBER OF SEQ ID NOS: 4096
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 413
/ LENGTH: 3469
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-104-047-413

Query Match          76.0%; Score 15.2; DB 3; Length 3469;
Best Local Similarity 85.0%; Pred. No. 2.1e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCAGGCCTT 20
   ||||| ||||| ||||| |||||
Db 3316 GTATCTGACCTTCAGGCCTT 3297

RESULT 29
US-08-578-096A-1
/ Sequence 1, Application US/08578096A
/ Patent No. 5980906
/ GENERAL INFORMATION:
/ APPLICANT:
/ TITLE OF INVENTION: Avian herpesvirus-based live recombinant
/ TITLE OF INVENTION: avian vaccine
/ NUMBER OF SEQUENCES: 28
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/578,096A
/ FILING DATE:
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 5838 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ ORIGINAL SOURCE:
/ ORGANISM: Herpesvirus of turkey
/ STRAIN: FC126
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 676..1209
/ OTHER INFORMATION: /function= "unknown"
/ OTHER INFORMATION: /product= "ORF1"
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: complement (1387..1941)
/ OTHER INFORMATION: /function= "unknown"
/ OTHER INFORMATION: /product= "ORF2"
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: complement (3573..5838)
/ OTHER INFORMATION: /function= "unknown"
/ OTHER INFORMATION: /product= "ORF3"
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 1403..1957
/ OTHER INFORMATION: /function= "unknown"
/ OTHER INFORMATION: /product= "ORF4"
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: complement (2287..3081)
/ OTHER INFORMATION: /function= "unknown"
/ OTHER INFORMATION: /product= "ORF5"
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: complement (1..479)
/ OTHER INFORMATION: /function= "unknown"
/ OTHER INFORMATION: /product= "unknown"
```

```
/ OTHER INFORMATION: /product= "ORF6"
US-08-578-096A-1

Query Match          76.0%; Score 15.2; DB 2; Length 5838;
Best Local Similarity 85.0%; Pred. No. 2.3e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCAGGCCTT 20
   ||||| ||||| ||||| |||||
Db 3025 GAAACAGACCTTCGGCCTT 3044

RESULT 30
US-09-240-426-1
/ Sequence 1, Application US/09240426
/ Patent No. 6045803
/ GENERAL INFORMATION:
/ APPLICANT:
/ TITLE OF INVENTION: Avian herpesvirus-based live recombinant
/ TITLE OF INVENTION: avian vaccine
/ NUMBER OF SEQUENCES: 28
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/240,426
/ FILING DATE:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/578,096
/ FILING DATE:
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 5838 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ ORIGINAL SOURCE:
/ ORGANISM: Herpesvirus of turkey
/ STRAIN: FC126
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 676..1209
/ OTHER INFORMATION: /function= "unknown"
/ OTHER INFORMATION: /product= "ORF1"
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: complement (1387..1941)
/ OTHER INFORMATION: /function= "unknown"
/ OTHER INFORMATION: /product= "ORF2"
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: complement (3573..5838)
/ OTHER INFORMATION: /function= "unknown"
/ OTHER INFORMATION: /product= "ORF3"
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 1403..1957
/ OTHER INFORMATION: /function= "unknown"
/ OTHER INFORMATION: /product= "ORF4"
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: complement (2287..3081)
/ OTHER INFORMATION: /function= "unknown"
/ OTHER INFORMATION: /product= "ORF5"
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: complement (1..479)
/ OTHER INFORMATION: /function= "unknown"
/ OTHER INFORMATION: /product= "ORF6"
US-09-240-426-1
```

```
Query Match          76.0%; Score 15.2; DB 3; Length 5838;
Best Local Similarity 85.0%; Pred. No. 2.3e+02; 7
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 GAAACTGACCTTCACGCCTT 20
      ||||| ||||| ||||| |||||
DB      3025 GAACAGACCTTCTCGGCTT 3044

RESULT 31
US-09-821-167-1
; Sequence 1, Application US/09821167
; Patent No. 6569668
; GENERAL INFORMATION:
; APPLICANT: Hosted Jr., Thomas J.
; APPLICANT: Horan, Ann C.
; TITLE OF INVENTION: Isolated Nucleic Acids from Micromonospora rosaria
; TITLE OF INVENTION: plasmid pMR2 and Vectors Made Therefrom
; FILE REFERENCE: IN011490
; CURRENT APPLICATION NUMBER: US/09/821,167
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: US 60/194,461
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 11188
; TYPE: DNA
; ORGANISM: Micromonospora rosaria
; FEATURE:
; NAME/KEY: RBS
; LOCATION: (6055)..(6059)
; NAME/KEY: RBS
; LOCATION: (6391)..(6394)
; NAME/KEY: RBS
; LOCATION: (8084)..(8088)
; NAME/KEY: RBS
; LOCATION: (9834)..(9837)
; NAME/KEY: RBS
; LOCATION: (10010)..(10012)
US-09-821-167-1

Query Match          76.0%; Score 15.2; DB 3; Length 11188;
Best Local Similarity 85.0%; Pred. No. 2.7e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 GAAACTGACCTTCACGCCTT 20
      ||||| ||||| ||||| |||||
DB      3753 GAACCTGACCGTCACGCCTT 3772

RESULT 32
US-09-949-016-11756
; Sequence 11756, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11756
; LENGTH: 11752
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(11865)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15297

Query Match          76.0%; Score 15.2; DB 3; Length 11865;
Best Local Similarity 85.0%; Pred. No. 2.7e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 GAAACTGACCTTCACGCCTT 20
      ||||| ||||| ||||| |||||
DB      1768 GAAGCGACCTTCACACCTT 1787

RESULT 34
US-09-949-016-15298
; Sequence 15298, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11756
; LENGTH: 11752
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(11865)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15297

Query Match          76.0%; Score 15.2; DB 3; Length 11865;
Best Local Similarity 85.0%; Pred. No. 2.7e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 GAAACTGACCTTCACGCCTT 20
      ||||| ||||| ||||| |||||
DB      1768 GAAGCGACCTTCACACCTT 1787

RESULT 34
US-09-949-016-15298
; Sequence 15298, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11756
; LENGTH: 11752
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(11865)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15297
```



```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15298
; LENGTH: 11865
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc.feature
; LOCATION: (1)...((11865)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15298

Query Match          76.0%; Score 15.2; DB 3; Length 11865;
Best Local Similarity 85.0%; Pred. No. 2.7e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACGCCTT 20
   ||| ||||| ||||| |||||
Db 1768 GAAGCGACCTTCACACCTT 1787

RESULT 35
US-09-949-016-13726
; Sequence 13726, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13726
; LENGTH: 23383
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13726

Query Match          76.0%; Score 15.2; DB 3; Length 23383;
Best Local Similarity 85.0%; Pred. No. 3.1e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAACTGACCTTCACGCCTT 20
   ||| ||||| ||||| |||||
Db 18864 GAACTGACATTCACCTCTT 18883

RESULT 36
US-09-949-016-16788/C
; Sequence 16788, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 16788
; LENGTH: 29905
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16788

Query Match          76.0%; Score 15.2; DB 3; Length 29905;
Best Local Similarity 85.0%; Pred. No. 3.2e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACGCCTT 20
   ||||| ||||| ||||| |||||
Db 10663 GAAACTCACCTTCCTCCTT 10644

RESULT 37
US-09-949-016-12985
; Sequence 12985, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12985
; LENGTH: 49164
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12985

Query Match          76.0%; Score 15.2; DB 3; Length 49164;
Best Local Similarity 85.0%; Pred. No. 3.6e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACGCCTT 20
   ||||| ||||| ||||| |||||
Db 39125 GAAACTGCCCTTCCTGCCTT 39144

RESULT 38
US-09-949-016-12986
; Sequence 12986, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12986
; LENGTH: 49164
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12986
```

```
Query Match          76.0%; Score 15.2; DB 3; Length 49164;
Best Local Similarity 85.0%; Pred. No. 3.6e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 GAAACTGACCTTCACGCCTT 20
      ||||| ||||| ||||| |||||
Db      39125 GAAACTGCCCTTCCTGCCTT 39144

Search completed: November 20, 2005, 22:01:48
Job time : 93.1111 secs

RESULT 39
US-09-949-016-13262
; Sequence 13262, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13262
; LENGTH: 49164
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13262

Query Match          76.0%; Score 15.2; DB 3; Length 49164;
Best Local Similarity 85.0%; Pred. No. 3.6e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 GAAACTGACCTTCACGCCTT 20
      ||||| ||||| ||||| |||||
Db      39125 GAAACTGCCCTTCCTGCCTT 39144

RESULT 40
US-09-949-016-13263
; Sequence 13263, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13263
; LENGTH: 49164
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13263

Query Match          76.0%; Score 15.2; DB 3; Length 49164;
Best Local Similarity 85.0%; Pred. No. 3.6e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2005, 18:34:25 ; Search time 602.469 Seconds  
(without alignments)  
274.516 Million cell updates/sec

Title: US-10-627-757-21

Perfect score: 20

Sequence: 1 gaaactgaccttcacgcctt 20

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA\_Main:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US11\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	20	7	US-10-627-757-21
2	20	100.0	1074	7	US-10-627-757-5
3	20	100.0	46951	6	US-10-091-281-2
c 4	16.8	84.0	891	6	US-10-369-493-28338
c 5	16.8	84.0	891	6	US-10-369-493-31096
c 6	16.8	84.0	2736	3	US-09-938-842A-235
c 7	16.8	84.0	2736	3	US-09-938-842A-235
c 8	16.8	84.0	3105	3	US-09-812-350-30
c 9	16.8	84.0	6376	3	US-09-812-350-31
c 10	16.4	82.0	3213	7	US-10-282-122A-42121
c 11	15.8	79.0	468	5	US-10-027-632-193257
c 12	15.8	79.0	468	6	US-10-027-632-193257
c 13	15.8	79.0	468	7	US-10-767-701-28120
c 14	15.8	79.0	510	3	US-09-854-867-370
c 15	15.8	79.0	510	9	US-10-786-970A-370
c 16	15.8	79.0	544	4	US-09-925-065A-935075
c 17	15.8	79.0	545	4	US-09-925-065A-939334
c 18	15.8	79.0	600	9	US-10-972-079-8953
c 19	15.8	79.0	633	4	US-09-925-065A-474965
c 20	15.8	79.0	641	8	US-10-653-047-595
c 21	15.8	79.0	654	5	US-10-027-632-106272
c 22	15.8	79.0	654	6	US-10-027-632-106272
c 23	15.8	79.0	661	4	US-09-925-065A-907420

c 24	15.8	79.0	661	4	US-09-925-065A-907421	Sequence 907421,
c 25	15.8	79.0	664	5	US-10-027-632-107122	Sequence 107122,
c 26	15.8	79.0	664	6	US-10-027-632-107122	Sequence 107122,
c 27	15.8	79.0	671	5	US-10-027-632-11769	Sequence 11769, A
c 28	15.8	79.0	671	6	US-10-027-632-11769	Sequence 11769, A
c 29	15.8	79.0	703	5	US-10-027-632-21598	Sequence 21598, A
c 30	15.8	79.0	703	6	US-10-027-632-21598	Sequence 21598, A
c 31	15.8	79.0	709	5	US-10-027-632-26559	Sequence 26559, A
c 32	15.8	79.0	709	6	US-10-027-632-26559	Sequence 26559, A
c 33	15.8	79.0	709	5	US-10-027-632-26559	Sequence 26559, A
c 34	15.8	79.0	709	6	US-10-027-632-26559	Sequence 26559, A
c 35	15.8	79.0	721	5	US-10-027-632-139475	Sequence 139475,
c 36	15.8	79.0	721	6	US-10-027-632-139475	Sequence 139475,
c 37	15.8	79.0	736	5	US-10-027-632-133909	Sequence 133909,
c 38	15.8	79.0	736	6	US-10-027-632-133909	Sequence 133909,
c 39	15.8	79.0	1166	4	US-09-925-065A-70040	Sequence 70040, A
c 40	15.8	79.0	1315	4	US-09-925-065A-551070	Sequence 551070,
c 41	15.8	79.0	1315	4	US-09-925-065A-551071	Sequence 551071,
c 42	15.8	79.0	1315	4	US-09-925-065A-551072	Sequence 551072,
c 43	15.8	79.0	1362	8	US-10-653-047-5689	Sequence 5689, Ap
c 44	15.8	79.0	1481	8	US-10-425-115-90258	Sequence 90258, A
c 45	15.8	79.0	1491	3	US-09-738-626-1380	Sequence 1380, Ap

ALIGNMENTS

RESULT 1  
US-10-627-757-21  
; Sequence 21, Application US/10627757  
; Publication No. US20040091914A1  
; GENERAL INFORMATION:  
; APPLICANT: MASASGO AKINORI  
; APPLICANT: KOUCHI YASUHIRO  
; APPLICANT: TAKAHATI TAKAYUKI  
; TITLE OF INVENTION: GENE ASSAY METHOD FOR PREDICTING GLAUCOMA ONSET RISK  
; FILE REFERENCE: Q76319  
; CURRENT APPLICATION NUMBER: US/10/627,757  
; CURRENT FILING DATE: 2003-07-28  
; PRIOR APPLICATION NUMBER: JP P2002-226612  
; PRIOR FILING DATE: 2002-08-02  
; NUMBER OF SEQ ID NOS: 40  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 21  
; LENGTH: 20  
; TYPE: DNA  
; ORGANISM: Artificial  
; FEATURE:  
; OTHER INFORMATION: Designed DNA based on OPTN gene  
US-10-627-757-21

Query Match	100.0%	Score 20;	DB 7;	Length 20;
Best Local Similarity	100.0%	Pred. No. 2.3;	0;	Gaps 0;
Matches	20;	Conservative 0;	Mismatches 0;	Indels 0;
Qy	1	GAAGTACCTTCACGCCTT 20		
Db	1	GAAGTACCTTCACGCCTT 20		
RESULT 2				
US-10-627-757-5				
; Sequence 5, Application US/10627757				
; Publication No. US20040091914A1				
; GENERAL INFORMATION:				
; APPLICANT: KOUCHI YASUHIRO				
; APPLICANT: MASASGO AKINORI				
; APPLICANT: TAKAHATI TAKAYUKI				
; TITLE OF INVENTION: GENE ASSAY METHOD FOR PREDICTING GLAUCOMA ONSET RISK				
; FILE REFERENCE: Q76319				
; CURRENT APPLICATION NUMBER: US/10/627,757				
; CURRENT FILING DATE: 2003-07-28				
; PRIOR APPLICATION NUMBER: JP P2002-226612				

```
; PRIOR FILING DATE: 2002-08-02
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
;   LENGTH: 1074
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-627-757-5

Query Match      100.0%; Score 20; DB 7; Length 1074;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GAAACTGACCTTCAGCCTT 20
      |||||
DB      261 GAAACTGACCTTCAGCCTT 280

RESULT 3
US-10-091-281-2
; Sequence 2, Application US/10091281
; Publication No. US20030190617A1
; GENERAL INFORMATION:
; APPLICANT: RAYMOND, VINCENT
; APPLICANT: SI, ERWIN
; APPLICANT: MORISETTE, JEAN
; TITLE OF INVENTION: OPTINEURIN NUCLEIC ACID MOLECULES AND USES THEREOF
; FILE REFERENCE: 13587.338
; CURRENT APPLICATION NUMBER: US/10/091,281
; CURRENT FILING DATE: 2002-03-06
; NUMBER OF SEQ ID NOS: 463
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 46951
;   TYPE: DNA
;   ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 391
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 691
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 709
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 887
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 894
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 987
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 1112
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 1505
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 1606
; OTHER INFORMATION: insertion of additional c residue
; NAME/KEY: allele
; LOCATION: 2405
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 2606
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 3313
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
```

```
; NAME/KEY: allele
; LOCATION: 3555
; OTHER INFORMATION: insertion of additional t residue
; NAME/KEY: allele
; LOCATION: 3625
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 3629
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 3882
; OTHER INFORMATION: insertion of additional t residue
; NAME/KEY: allele
; LOCATION: 3988
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 4452
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: repeat region
; LOCATION: 598..878
; OTHER INFORMATION: repeat element
; NAME/KEY: repeat region
; LOCATION: 938..957
; OTHER INFORMATION: Short repeat element
; NAME/KEY: repeat region
; LOCATION: 1002..1329
; OTHER INFORMATION: ALU repeat element
; NAME/KEY: repeat region
; LOCATION: 2288..2587
; OTHER INFORMATION: ALU repeat element
; NAME/KEY: misc feature
; LOCATION: 5054
; OTHER INFORMATION: putative transcription start site
US-10-091-281-2

Query Match      100.0%; Score 20; DB 6; Length 46951;
Best Local Similarity 100.0%; Pred. No. 3.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GAAACTGACCTTCAGCCTT 20
      |||||
DB      20871 GAAACTGACCTTCAGCCTT 20890

RESULT 4
US-10-369-493-28338/c
; Sequence 28338, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 28338
; LENGTH: 891
; TYPE: DNA
; ORGANISM: Burkholderia fungorum
US-10-369-493-28338

Query Match      84.0%; Score 16.8; DB 6; Length 891;
Best Local Similarity 90.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 GAAACTGACCTTCAGCCTT 20
```

Db 465 GAAACTGATTTTCACGCCTT 446  
|||||

RESULT 5  
US-10-369-493-31096/c  
; Sequence 31096, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10 (52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 31096  
; LENGTH: 891  
; TYPE: DNA  
; ORGANISM: Burkholderia cepacia  
US-10-369-493-31096

Query Match 84.0%; Score 16.8; DB 6; Length 891;  
Best Local Similarity 90.0%; Pred. No. 1.4e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACGCCTT 20  
|||||  
Db 465 GAAACTGATTTTCACGCCTT 446  
|||||

RESULT 6  
US-09-938-842A-235  
; Sequence 235, Application US/09938842A  
; Patent No. US20020160378A1  
; GENERAL INFORMATION:  
; APPLICANT: Harper, Jeff  
; APPLICANT: Kreps, Joel  
; APPLICANT: Wang, Xun  
; APPLICANT: Zhu, Tong  
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
; TITLE OF INVENTION: SAME, AND METHODS OF USE  
; FILE REFERENCE: SCRIPI300-3  
; CURRENT APPLICATION NUMBER: US/09/938,842A  
; CURRENT FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: US 60/227,866  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: US 60/264,647  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/300,111  
; PRIOR FILING DATE: 2001-06-22  
; NUMBER OF SEQ ID NOS: 5379  
; SEQ ID NO 235  
; LENGTH: 2736  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-09-938-842A-235

Query Match 84.0%; Score 16.8; DB 3; Length 2736;  
Best Local Similarity 90.0%; Pred. No. 1.5e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACGCCTT 20  
|||||  
Db 1270 GAAATTGAATTCACGCCTT 1289  
|||||

RESULT 7  
US-09-938-842A-235  
; Sequence 235, Application US/09938842A  
; Publication No. US20040009476A9  
; GENERAL INFORMATION:  
; APPLICANT: Harper, Jeff  
; APPLICANT: Kreps, Joel  
; APPLICANT: Wang, Xun  
; APPLICANT: Zhu, Tong  
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
; TITLE OF INVENTION: SAME, AND METHODS OF USE  
; FILE REFERENCE: SCRIPI300-3  
; CURRENT APPLICATION NUMBER: US/09/938,842A  
; CURRENT FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: US 60/227,866  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: US 60/264,647  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/300,111  
; PRIOR FILING DATE: 2001-06-22  
; NUMBER OF SEQ ID NOS: 5379  
; SEQ ID NO 235  
; LENGTH: 2736  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-09-938-842A-235

Query Match 84.0%; Score 16.8; DB 3; Length 2736;  
Best Local Similarity 90.0%; Pred. No. 1.5e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACGCCTT 20  
|||||  
Db 1270 GAAATTGAATTCACGCCTT 1289  
|||||

RESULT 8  
US-09-812-350-30  
; Sequence 30, Application US/09812350  
; Publication No. US20020053097A1  
; GENERAL INFORMATION:  
; APPLICANT: Lindquist, Susan  
; APPLICANT: Queitsch, Christine  
; APPLICANT: Vierling, Elizabeth  
; TITLE OF INVENTION: Transgenic Plants Containing Heat Shock Protein  
; FILE REFERENCE: P01979US2  
; CURRENT APPLICATION NUMBER: US/09/812,350  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/190,769  
; PRIOR FILING DATE: 2000-03-20  
; PRIOR APPLICATION NUMBER: US 60/198,116  
; PRIOR FILING DATE: 2000-04-18  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 30  
; LENGTH: 3105  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-09-812-350-30

Query Match 84.0%; Score 16.8; DB 3; Length 3105;  
Best Local Similarity 90.0%; Pred. No. 1.5e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACGCCTT 20  
|||||  
Db 1432 GAAATTGAATTCACGCCTT 1451  
|||||

RESULT 9:  
US-09-812-350-31  
; Sequence 31, Application US/09812350  
; Publication No. US20020053097A1

```

; GENERAL INFORMATION:
; APPLICANT: Lindquist, Susan
; APPLICANT: Queitsch, Christine
; APPLICANT: Vierling, Elizabeth
; TITLE OF INVENTION: Transgenic Plants Containing Heat Shock Protein
; FILE REFERENCE: P01979US2
; CURRENT APPLICATION NUMBER: US/09/812,350
; CURRENT FILING DATE: 2001-03-20
; PRIOR FILING DATE: 2000-03-20
; PRIOR APPLICATION NUMBER: US 60/190,769
; PRIOR FILING DATE: 2000-03-20
; PRIOR APPLICATION NUMBER: US 60/198,116
; PRIOR FILING DATE: 2000-04-18
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 31
; LENGTH: 6376
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; US-09-812-350-31

Query Match      84.0%; Score 16.8; DB 3; Length 6376;
Best Local Similarity 90.0%; Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1  GAAACTGACCTTCACGCCTT 20
        ||||| ||||| ||||| |||||
Db      2485  GAAATTGAACTTCACGCCTT 2504

RESULT 10
US-10-282-122A-42121/c
; Sequence 42121, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1

```



; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 193257  
; LENGTH: 468  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-193257

Query Match 79.0%; Score 15.8; DB 6; Length 468;  
Best Local Similarity 89.5%; Pred. No. 4.6e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACGCCT 19  
|||||  
Db 464 GAAACTGACCTTCACGCCT 446

RESULT 13  
US-10-767-701-28120  
; Sequence 28120, Application US/10767701  
; Publication No. US20040172684A1  
; GENERAL INFORMATION:  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53535)B  
; CURRENT APPLICATION NUMBER: US/10767,701  
; CURRENT FILING DATE: 2004-01-29  
; NUMBER OF SEQ ID NOS: 63128  
; SEQ ID NO 28120  
; LENGTH: 468  
; TYPE: DNA  
; ORGANISM: Sorghum bicolor  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 7535528  
US-10-767-701-28120

Query Match 79.0%; Score 15.8; DB 7; Length 468;  
Best Local Similarity 89.5%; Pred. No. 4.6e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAACTGACCTTCACGCCTT 20  
|||||  
Db 198 AACTGACCTTCACGCCGT 216

RESULT 14  
US-09-854-867-370  
; Sequence 370, Application US/09854867  
; Publication No. US20030224356A1  
; GENERAL INFORMATION:  
; APPLICANT: JOAN, KNOLL H  
; APPLICANT: ROGAN, PETER K  
; TITLE OF INVENTION: SINGLE COPY GENOMIC HYBRIDIZATION PROBES AND METHOD OF GENERATING  
; FILE REFERENCE: 30307  
; CURRENT APPLICATION NUMBER: US/09/854,867  
; CURRENT FILING DATE: 2003-05-08  
; NUMBER OF SEQ ID NOS: 613  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 370  
; LENGTH: 510  
; TYPE: DNA

; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: repeat\_region  
; LOCATION: (1)..(510)  
; OTHER INFORMATION: ltr10b  
US-09-854-867-370

Query Match 79.0%; Score 15.8; DB 3; Length 510;  
Best Local Similarity 89.5%; Pred. No. 4.6e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAACGTGACCTTCACGCCTT 20  
|||||  
Db 217 AAACGTGATCTTCCCGCCTT 235

RESULT 15  
US-10-786-970A-370  
; Sequence 370, Application US/10786970A  
; Publication No. US2005006449A1  
; GENERAL INFORMATION:  
; APPLICANT: JOAN, KNOLL  
; APPLICANT: ROGAN, PETER  
; TITLE OF INVENTION: SINGLE COPY GENOMIC HYBRIDIZATION PROBES AND METHOD OF GENERATING  
; FILE REFERENCE: 30307  
; CURRENT APPLICATION NUMBER: US/10/786,970A  
; CURRENT FILING DATE: 2004-02-24  
; PRIOR APPLICATION NUMBER: US/09/573,080  
; PRIOR FILING DATE: 2000-05-16  
; NUMBER OF SEQ ID NOS: 479  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 370  
; LENGTH: 510  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: repeat\_region  
; LOCATION: (1)..(510)  
; OTHER INFORMATION: ltr10b  
; PUBLICATION INFORMATION:  
; AUTHORS: Jurka, J; Walichewicz, J; Milosavljevic, A  
; TITLE: Prototypic sequences for human repetitive DNA  
; JOURNAL: Journal of Molecular Evolution  
; VOLUME: 35  
; ISSUE: 4  
; PAGES: 286-291  
; DATE: 1992-10-  
; DATABASE ACCESSION NUMBER: Database of repetitive elements (repbase)  
; DATABASE ENTRY DATE:  
; DATABASE ENTRY DATE: 1996-01-26  
US-10-786-970A-370

Query Match 79.0%; Score 15.8; DB 9; Length 510;  
Best Local Similarity 89.5%; Pred. No. 4.6e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAACGTGACCTTCACGCCTT 20  
|||||  
Db 217 AAACGTGATCTTCCCGCCTT 235

RESULT 16  
US-09-925-065A-935075/c  
; Sequence 935075, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08

; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 935075  
; LENGTH: 544  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-935075

Query Match 79.0%; Score 15.8; DB 4; Length 544;  
Best Local Similarity 89.5%; Pred. No. 4.6e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCAGCCT 19  
||||| ||||| ||||| |||||  
Db 490 GAAATGACCTTCAGCCT 472

RESULT 17  
US-09-925-065A-939334  
; Sequence 939334, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; FILE REFERENCE: 108827,135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 939334  
; LENGTH: 545  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-939334

Query Match 79.0%; Score 15.8; DB 4; Length 545;  
Best Local Similarity 89.5%; Pred. No. 4.6e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCAGCCT 19  
||||| ||||| ||||| |||||  
Db 60 GAAATGACCTTCAGCCT 78

RESULT 18  
US-10-972-079-8953/c  
; Sequence 8953, Application US/10972079  
; Publication No. US20050153317A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: Denise, Sue K.  
; APPLICANT: ROSENFELD, David

; APPLICANT: KERR, Richard  
; APPLICANT: BATES, Stephen  
; APPLICANT: HOLM, Tom  
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEEF  
; FILE REFERENCE: MM1110-2  
; CURRENT APPLICATION NUMBER: US/10/972,079  
; CURRENT FILING DATE: 2004-10-22  
; PRIOR APPLICATION NUMBER: US 60/514,333  
; PRIOR FILING DATE: 2003-10-24  
; NUMBER OF SEQ ID NOS: 96631  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 8953  
; LENGTH: 600  
; TYPE: DNA  
; ORGANISM: Chicken 19866894204079\_1  
US-10-972-079-8953

Query Match 79.0%; Score 15.8; DB 9; Length 600;  
Best Local Similarity 89.5%; Pred. No. 4.6e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAACTGACCTTCAGCCT 19  
||||| ||||| ||||| |||||  
Db 30 GAACTGACCTTCACATCT 12

RESULT 19  
US-09-925-065A-474965/c  
; Sequence 474965, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; FILE REFERENCE: 108827,135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 474965  
; LENGTH: 633  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-474965

Query Match 79.0%; Score 15.8; DB 4; Length 633;  
Best Local Similarity 89.5%; Pred. No. 4.6e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAACCTGACCTTCAGCCTT 20  
||||| ||||| ||||| |||||  
Db 338 AACCTGACCTTCAGCCTT 320

RESULT 20  
US-10-653-047-595  
; Sequence 595, Application US/10653047  
; Publication No. US20040229367A1  
; GENERAL INFORMATION:  
; APPLICANT: Randy M. Berka  
; APPLICANT: Michael W. Rey  
; APPLICANT: Jeffrey R. Shuster

```

? APPLICANT: Sakari Kauppinen
? APPLICANT: Ib Groth Clausen
? APPLICANT: Peter Bjarke Olsen
? TITLE OF INVENTION: Methods For Monitoring Multiple Gene
? TITLE OF INVENTION: Expression
? FILE REFERENCE: 5849.200-US
? CURRENT APPLICATION NUMBER: US/10/653,047
? CURRENT FILING DATE: 2003-08-29
? PRIOR APPLICATION NUMBER: US/09/533,559
? PRIOR FILING DATE: 2000-03-22
? PRIOR APPLICATION NUMBER: 09/273,623
? PRIOR FILING DATE: 1999-03-22
? NUMBER OF SEQ ID NOS: 7860
? SOFTWARE: Fast-SEQ for Windows Version 4.0
? SEQ ID NO 595
? LENGTH: 641
? TYPE: DNA
? ORGANISM: Fusarium venenatum
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (1)...(641)
? OTHER INFORMATION: n = A, T, C or G
? US-10-653-047-595

```

Query Match 79.0%; Score 15.8; DB 8; Length 641;  
Best Local Similarity 89.5%; Pred. NO. 4.7e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GAAACTGACCTTCACGCCT 19  
|||  
Db 469 GAAACTGTCCTTCACGACT 487

```

RESULT 21
US-10-027-632-106272
; Sequence 106272, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome

```

```

, REFERENCE: 10062115
, CURRENT APPLICATION NUMBER: US/10/027,632
, CURRENT FILING DATE: 2002-04-30
, PRIOR APPLICATION NUMBER: US 60/218,006
, PRIOR FILING DATE: 2000-07-12
, PRIOR APPLICATION NUMBER: US 60/198,676
, PRIOR FILING DATE: 2000-04-20
, PRIOR APPLICATION NUMBER: US 60/193,483
, PRIOR FILING DATE: 2000-03-29
, PRIOR APPLICATION NUMBER: US 60/185,218
, PRIOR FILING DATE: 2000-02-24
, PRIOR APPLICATION NUMBER: US 60/167,363
, PRIOR FILING DATE: 1999-11-23
, PRIOR APPLICATION NUMBER: US 60/156,358
, PRIOR FILING DATE: 1999-09-28
, PRIOR APPLICATION NUMBER: US 60/146,002
, PRIOR FILING DATE: 1999-08-09
, NUMBER OF SEQ ID NOS: 325720
, SOFTWARE: FastSeq for Windows Version 4.0
, SEQ ID NO 106272
, LENGTH: 654
, TYPE: DNA
, ORGANISM: Human
US-10-027-632-106272

```

```

Query Match          79.0%; Score 15.8; DB 5; Length 654;
Best Local Similarity 89.5%; Pred. No. 4.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

Qy 2 A A A C T G A C C T T C A C G C C T T 20  
|||  
pb 48 A C A C T G A G C C T T C A C G C C T T 66

```

RESULT 22
US-10-027-632-106272
; Sequence 106272, Application US/10027632
; Publication NO. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 106272
; LENGTH: 654
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-106272

```

```
Query Match          79.0%; Score 15.8; DB 6; Length 654;
Best Local Similarity 89.5%; Pred. No. 4.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

Qy 2 A A A C T G A C C T T C A C G C C T T 20  
| | | | | | | | | | | | | | | | | |  
Db 48 A C A C T G A G C T T C A C G C C T T 66

```

RESULT 23
US-09-925-065A-907420/c
; Sequence 907420, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925, 065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243, 096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 907420
; LENGTH: 661
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-907420

```

Query Match 79.0%; Score 15.8; DB 4; Length 661;  
Best Local Similarity 89.5%; Pred. No. 4.7e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GAACTGACCTTCACGCCT 19  
| | | | | | | | | | | | | | | | | | | | |  
Db 197 GAACTGAGCTTCATGCCT 179

RESULT 24  
US-09-925-065A-907421/c  
; Sequence 907421, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; PRIOR FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 907421  
; LENGTH: 661  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-907421

Query Match 79.0%; Score 15.8; DB 4; Length 661;  
Best Local Similarity 89.5%; Pred. No. 4.7e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GAACTGACCTTCACGCCT 19  
| | | | | | | | | | | | | | | | | | | | |  
Db 197 GAACTGAGCTTCATGCCT 179

RESULT 25  
US-10-027-632-107122  
; Sequence 107122, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; PRIOR FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720

; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 107122  
; LENGTH: 664  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-107122

Query Match 79.0%; Score 15.8; DB 5; Length 664;  
Best Local Similarity 89.5%; Pred. No. 4.7e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AAACGTGACCTTCACGCCTT 20  
| | | | | | | | | | | | | | | | | | | | |  
Db 295 ACACGTGAGCTTCACGCCTT 313

RESULT 26  
US-10-027-632-107122  
; Sequence 107122, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; PRIOR FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 107122  
; LENGTH: 664  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-107122

Query Match 79.0%; Score 15.8; DB 6; Length 664;  
Best Local Similarity 89.5%; Pred. No. 4.7e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AAACGTGACCTTCACGCCTT 20  
| | | | | | | | | | | | | | | | | | | | |  
Db 295 ACACGTGAGCTTCACGCCTT 313

RESULT 27  
US-10-027-632-11769/c  
; Sequence 11769, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; PRIOR FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20

```
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11769
; LENGTH: 671
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-11769
```

```
Query Match          79.0%; Score 15.8; DB 5; Length 671;
Best Local Similarity 89.5%; Pred. No. 4.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      1  GAAACTGACCTTCACGCCT 19
          ||||| ||||| ||||| |||||
DB      540  GAAACTGACCATCAGGCCT 522
```

## RESULT 28

```
US-10-027-632-11769/c
; Sequence 11769, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
```

```
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11769
; LENGTH: 671
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-11769
```

```
Query Match          79.0%; Score 15.8; DB 6; Length 671;
Best Local Similarity 89.5%; Pred. No. 4.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      1  GAAACTGACCTTCACGCCT 19
          ||||| ||||| ||||| |||||
DB      540  GAAACTGACCATCAGGCCT 522
```

## RESULT 29

```
US-10-027-632-21598/c
; Sequence 21598, Application US/10027632
; Publication No. US20020198371A1
```

## ; GENERAL INFORMATION:

```
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21598
; LENGTH: 703
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-21598
```

```
Query Match          79.0%; Score 15.8; DB 5; Length 703;
Best Local Similarity 89.5%; Pred. No. 4.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      2  AAACGTGACCTTCACGCCTT 20
          ||||| ||||| ||||| |||||
DB      606  ACACGTGACCTTCACGCCTT 588
```

## RESULT 30

```
US-10-027-632-21598/c
; Sequence 21598, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
```

```
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21598
; LENGTH: 703
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-21598
```

```
Query Match          79.0%; Score 15.8; DB 6; Length 703;
Best Local Similarity 89.5%; Pred. No. 4.7e+02;
```

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AAACGTGACCTTCACGCCTT 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 606 ACACTGAGCTTCACGCCTT 588

## RESULT 31

US-10-027-632-26559  
; Sequence 26559, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US 10/027,632  
; PRIOR FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 26559  
; LENGTH: 709  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-26559

Query Match 79.0%; Score 15.8; DB 5; Length 709;

Best Local Similarity 89.5%; Pred. No. 4.7e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AAACGTGACCTTCACGCCTT 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 290 ACACTGAGCTTCACGCCTT 308

## RESULT 32

US-10-027-632-26560  
; Sequence 26560, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US 10/027,632  
; PRIOR FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002

; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 26560  
; LENGTH: 709  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-26560

Query Match 79.0%; Score 15.8; DB 5; Length 709;

Best Local Similarity 89.5%; Pred. No. 4.7e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AAACGTGACCTTCACGCCTT 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 290 ACACTGAGCTTCACGCCTT 308

## RESULT 33

US-10-027-632-26559  
; Sequence 26559, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US 10/027,632  
; PRIOR FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 26559  
; LENGTH: 709  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-26559

Query Match 79.0%; Score 15.8; DB 6; Length 709;

Best Local Similarity 89.5%; Pred. No. 4.7e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AAACGTGACCTTCACGCCTT 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 290 ACACTGAGCTTCACGCCTT 308

## RESULT 34

US-10-027-632-26560  
; Sequence 26560, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US 10/027,632  
; PRIOR FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 26560  
; LENGTH: 709  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-26560

Query Match 79.0%; Score 15.8; DB 6; Length 709;  
Best Local Similarity 89.5%; Pred. No. 4.7e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAACGACCTTCACGCCTT 20  
| | | | | | | | | | | | | | | | | | | | | |  
DB 290 ACACTGAGCTTCACGCCTT 308

RESULT 35  
US-10-027-632-139475  
; Sequence 139475, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 139475  
; LENGTH: 721  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-139475

Query Match 79.0%; Score 15.8; DB 5; Length 721;  
Best Local Similarity 89.5%; Pred. No. 4.7e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAACGACCTTCACGCCTT 20  
| | | | | | | | | | | | | | | | | | | | | |  
DB 452 AAACGACCTTCACGCCTT 470

RESULT 36  
US-10-027-632-139475

; Sequence 139475, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 139475  
; LENGTH: 721  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-139475

Query Match 79.0%; Score 15.8; DB 6; Length 721;  
Best Local Similarity 89.5%; Pred. No. 4.7e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAACGACCTTCACGCCTT 20  
| | | | | | | | | | | | | | | | | | | | | |  
DB 452 AAACGACCTTCACGCCTT 470

RESULT 37  
US-10-027-632-133909  
; Sequence 133909, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 133909  
; LENGTH: 736  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-133909

Query Match 79.0%; Score 15.8; DB 5; Length 736;  
Best Local Similarity 89.5%; Pred. No. 4.7e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCAGGCCT 19  
||| ||||| |||||  
DB 446 GAAATGACCTTCAGGCCT 464

RESULT 38  
US-10-027-632-133909  
; Sequence 133909, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 133909  
; LENGTH: 736  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-133909

Query Match 79.0%; Score 15.8; DB 6; Length 736;  
Best Local Similarity 89.5%; Pred. No. 4.7e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCAGGCCT 19  
||| ||||| |||||  
DB 446 GAAATGACCTTCAGGCCT 464

RESULT 39  
US-09-925-065A-70040  
; Sequence 70040, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086

; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 70040  
; LENGTH: 1166  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-70040

Query Match 79.0%; Score 15.8; DB 4; Length 1166;  
Best Local Similarity 89.5%; Pred. No. 4.8e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAACGTGACCTTCAGGCCTT 20  
||| ||||| |||||  
DB 392 AAAATGACCTTCAGGCCTT 410

RESULT 40  
US-09-925-065A-551070  
; Sequence 551070, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 551070  
; LENGTH: 1315  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-551070

Query Match 79.0%; Score 15.8; DB 4; Length 1315;  
Best Local Similarity 89.5%; Pred. No. 4.9e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAACGTGACCTTCAGGCCTT 20  
||| ||||| |||||  
DB 392 AAAATGACCTTCAGGCCTT 410

Search completed: November 21, 2005, 05:03:37  
Job time : 603.469 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2005, 21:55:46 ; Search time 279.506 Seconds  
(without alignments)  
9.405 Million cell updates/sec

Title: US-10-627-757-21

Perfect score: 20

Sequence: 1 gaaactgaccttcacgcctt 20

Scoring table:

IDENTITY NUC

Gap 10.0 , Gapext 1.0

Searched: 3196817 seqs, 65720914 residues

Total number of hits satisfying chosen parameters: 6393634

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_NA\_New.\*

- 1: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
- 2: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
- 5: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*
- 7: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*
- 9: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*
- 10: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14.8	74.0	2313	9 US-11-082-389-357	Sequence 357, App
2	13.8	69.0	19	8 US-11-101-244-1313009	Sequence 1313009,
3	13.8	69.0	19	9 US-11-083-784-1313009	Sequence 1313009,
4	13.6	68.0	1725	9 US-11-082-389-347	Sequence 347, App
5	13.4	67.0	2623	1 US-10-131-826A-63	Sequence 63, App
6	13.4	67.0	2600	1 US-10-949-720-391	Sequence 391, App
7	13.2	66.0	19	8 US-11-101-244-25078	Sequence 25078, A
8	13.2	66.0	19	8 US-11-101-244-270793	Sequence 270793,
9	13.2	66.0	19	8 US-11-101-244-270848	Sequence 270848,
10	13.2	66.0	19	8 US-11-101-244-362349	Sequence 362349,
11	13.2	66.0	19	8 US-11-101-244-1212199	Sequence 1212199,
12	13.2	66.0	19	8 US-11-101-244-1502436	Sequence 1502436,
13	13.2	66.0	19	8 US-11-083-784-25078	Sequence 25078, A
14	13.2	66.0	19	9 US-11-083-784-270793	Sequence 270793,
15	13.2	66.0	19	9 US-11-083-784-270848	Sequence 270848,
16	13.2	66.0	19	9 US-11-083-784-362349	Sequence 362349,
17	13.2	66.0	19	9 US-11-083-784-1212199	Sequence 1212199,
18	13.2	66.0	19	9 US-11-083-784-1502436	Sequence 1502436,
19	12.8	64.0	19	8 US-11-101-244-25073	Sequence 25073, A
20	12.8	64.0	19	8 US-11-101-244-25117	Sequence 25117, A
21	12.8	64.0	19	8 US-11-101-244-270762	Sequence 270762,
22	12.8	64.0	19	8 US-11-101-244-270835	Sequence 270835,
23	12.8	64.0	19	8 US-11-101-244-270911	Sequence 270911,

c	24	12.8	64.0	19	8	US-11-101-244-270975	Sequence 270975,
	25	12.8	64.0	19	8	US-11-101-244-466224	Sequence 466224,
	26	12.8	64.0	19	8	US-11-101-244-940667	Sequence 940667,
	27	12.8	64.0	19	8	US-11-101-244-1114497	Sequence 1114497,
	28	12.8	64.0	19	9	US-11-083-784-25073	Sequence 25073, A
	29	12.8	64.0	19	9	US-11-083-784-25117	Sequence 25117, A
c	30	12.8	64.0	19	9	US-11-083-784-270762	Sequence 270762,
c	31	12.8	64.0	19	9	US-11-083-784-270835	Sequence 270835,
c	32	12.8	64.0	19	9	US-11-083-784-270911	Sequence 270911,
c	33	12.8	64.0	19	9	US-11-083-784-270975	Sequence 270975,
	34	12.8	64.0	19	9	US-11-083-784-466224	Sequence 466224,
	35	12.8	64.0	19	9	US-11-083-784-940667	Sequence 940667,
	36	12.8	64.0	19	9	US-11-083-784-1114497	Sequence 1114497,
	37	12.6	63.0	19	8	US-11-101-244-337258	Sequence 337258,
	38	12.6	63.0	19	8	US-11-101-244-355617	Sequence 355617,
	39	12.6	63.0	19	8	US-11-101-244-452754	Sequence 452754,
c	40	12.6	63.0	19	8	US-11-101-244-634465	Sequence 634465,
	41	12.6	63.0	19	8	US-11-101-244-948971	Sequence 948971,
	42	12.6	63.0	19	8	US-11-101-244-1088207	Sequence 1088207,
	43	12.6	63.0	19	8	US-11-101-244-1147318	Sequence 1147318,
	44	12.6	63.0	19	8	US-11-101-244-1187685	Sequence 1187685,
	45	12.6	63.0	19	8	US-11-101-244-1258436	Sequence 1258436,

#### ALIGNMENTS

##### RESULT 1

US-11-082-389-357  
; Sequence 357, Application US/11082389  
; Publication No. US20050244935A1  
; GENERAL INFORMATION:  
; APPLICANT: Pompejus, Markus  
; APPLICANT: Kroger, Burkhard  
; APPLICANT: Schroder, Hartwig  
; APPLICANT: Zeider, Oskar  
; APPLICANT: Haberhauer, Gregor  
; TITLE OF INVENTION: INVOLVED IN MEMBRANE SYNTHESIS AND MEMBRANE  
; TITLE OF INVENTION: TRANSPORT  
; FILE REFERENCE: BGI-131CPCN  
; CURRENT APPLICATION NUMBER: US/11082,389  
; CURRENT FILING DATE: 2005-03-16  
; PRIOR APPLICATION NUMBER: US 09/603024  
; PRIOR FILING DATE: 2000-06-23  
; PRIOR APPLICATION NUMBER: US 60/141031  
; PRIOR FILING DATE: 1999-06-25  
; PRIOR APPLICATION NUMBER: US 60/143262  
; PRIOR FILING DATE: 1999-07-09  
; PRIOR APPLICATION NUMBER: US 60/151281  
; PRIOR FILING DATE: 1999-08-27  
; PRIOR APPLICATION NUMBER: DE 19930487.4  
; PRIOR FILING DATE: 1999-07-01  
; PRIOR APPLICATION NUMBER: DE 19930489.0  
; PRIOR FILING DATE: 1999-07-01  
; PRIOR APPLICATION NUMBER: DE 19931549.3  
; PRIOR FILING DATE: 1999-07-08  
; PRIOR APPLICATION NUMBER: DE 19931550.7  
; PRIOR FILING DATE: 1999-07-08  
; PRIOR APPLICATION NUMBER: DE 19932134.5  
; PRIOR FILING DATE: 1999-07-09  
; PRIOR APPLICATION NUMBER: DE 19941379.7  
; PRIOR FILING DATE: 1999-08-31  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 446  
; SEQ ID NO 357  
; LENGTH: 2313  
; TYPE: DNA  
; ORGANISM: Corynebacterium glutamicum  
; FEATURE: CDS  
; NAME/KEY: CDS  
; LOCATION: (101)..(2290)  
; OTHER INFORMATION: RXS00479



```

; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P33301C128
; CURRENT APPLICATION NUMBER: US/10/131,826A
; CURRENT FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 63
; LENGTH: 2623
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-131-826A-63

Query Match 67.0%; Score 13.4; DB 1; Length 2623;
Best Local Similarity 93.3%; Pred. No. 4.9e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 ACTGACCTTCAGGCC 18
|||:|||||
Db 1523 ACTGACCTTCATCC 1537

RESULT 6
US-10-949-720-391
; Sequence 391, Application US/10949720
; Publication No. US20050249736A1
; GENERAL INFORMATION:
; APPLICANT: Krasnoperov, Valery
; APPLICANT: Zozulya, Sergey
; APPLICANT: Kertes, Nathalie
; APPLICANT: Reddy, Ramachandra
; APPLICANT: Gill, Parkash
; TITLE OF INVENTION: POLYPEPTIDE COMPOUNDS FOR INHIBITING
; TITLE OF INVENTION: ANGIOGENESIS AND TUMOR GROWTH
; FILE REFERENCE: VASQ-P02-002

; GENERAL INFORMATION:
; CURRENT APPLICATION NUMBER: US/10/949,720
; CURRENT FILING DATE: 2004-09-23
; PRIOR APPLICATION NUMBER: US 60/454,432
; PRIOR FILING DATE: 2003-03-12
; PRIOR APPLICATION NUMBER: US 60/454,300
; PRIOR FILING DATE: 2003-03-12
; PRIOR APPLICATION NUMBER: US 10/800,350
; PRIOR FILING DATE: 2004-03-12
; NUMBER OF SEQ ID NOS: 425
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 391
; LENGTH: 26000
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-949-720-391

Query Match 67.0%; Score 13.4; DB 1; Length 26000;
Best Local Similarity 93.3%; Pred. No. 3.9e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCAC 15
|||:|||||
Db 2061 GAAACTGACCTACAC 2075

RESULT 7
US-11-101-244-25078
; Sequence 25078, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 25078
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-25078

Query Match 66.0%; Score 13.2; DB 8; Length 19;
Best Local Similarity 66.7%; Pred. No. 8.6e+02;
Matches 12; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCAGCC 18
|||:|||||
Db 2 GAAACUGCCCUCAUGAC 19

RESULT 8
US-11-101-244-270793/c
; Sequence 270793, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
```

```
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 270793
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-270793
```

```
Query Match          66.0%; Score 13.2; DB 8; Length 19;
Best Local Similarity 83.3%; Pred. No. 8.6e+02;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      2 AAACGTGACCTTCACGCCT 19
        |||||
Db      19 AAACGTGACCTTCTGGCTT 2
```

## RESULT 9

```
US-11-101-244-270848/c
; Sequence 270848, Application US/11/101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 270848
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-270848
```

```
Query Match          66.0%; Score 13.2; DB 8; Length 19;
Best Local Similarity 83.3%; Pred. No. 8.6e+02;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      2 AAACGTGACCTTCACGCCT 19
        |||||
Db      19 AAACGTGACCTTCTGGCTT 2
```

## RESULT 10

```
US-11-101-244-362349
; Sequence 362349, Application US/11/101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
```

```
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 362349
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-362349
```

```
Query Match          66.0%; Score 13.2; DB 8; Length 19;
Best Local Similarity 61.1%; Pred. No. 8.6e+02;
Matches 11; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      2 AAACGTGACCTTCACGCCT 19
        |||||
Db      1 AAACUGACCUUGACUACU 18
```

## RESULT 11

```
US-11-101-244-1212199
; Sequence 1212199, Application US/11/101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 1212199
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-1212199
```

```
Query Match          66.0%; Score 13.2; DB 8; Length 19;
Best Local Similarity 61.1%; Pred. No. 8.6e+02;
Matches 11; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      3 AACTGACCTTCACGCCTT 20
        |||||
Db      2 AACUGAUCUCCACUCCUU 19
```

## RESULT 12

```
US-11-101-244-1502436
; Sequence 1502436, Application US/11/101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
```

```
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 1502436
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-1502436
```

```
Query Match      66.0%; Score 13.2; DB 8; Length 19;
Best Local Similarity 61.1%; Pred. No. 8.6e+02;
Matches 11; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 3 AACTGACCTTCACGCCTT 20
    |||||:|||||:|:|
Db 2 AACAGAGCUUACACCUU 19
```

```
RESULT 13
US-11-083-784-25078
; Sequence 25078, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR APPLICATION NUMBER: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 25078
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-25078
```

```
Query Match      66.0%; Score 13.2; DB 9; Length 19;
Best Local Similarity 66.7%; Pred. No. 8.6e+02;
Matches 12; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 GAAACTGACCTTCACGCC 18
    |||||:|||||:|
Db 2 GAAACUGCCCUUACUAGAC 19
```

```
RESULT 14
US-11-083-784-270793/c
; Sequence 270793, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
```

```
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR APPLICATION NUMBER: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 270793
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-270793
```

```
Query Match      66.0%; Score 13.2; DB 9; Length 19;
Best Local Similarity 83.3%; Pred. No. 8.6e+02;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 2 AAACGTGACCTTCACGCCT 19
    |||||:|||||:|
Db 19 AAACGTGACCTTCTGGCTT 2
```

```
RESULT 15
US-11-083-784-270848/c
; Sequence 270848, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR APPLICATION NUMBER: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 270848
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-270848
```

```
Query Match      66.0%; Score 13.2; DB 9; Length 19;
Best Local Similarity 83.3%; Pred. No. 8.6e+02;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 2 AAACGTGACCTTCACGCCT 19
    |||||:|||||:|
Db 19 AAACGTGACCTTCTGGCTT 2
```

```
RESULT 16
US-11-083-784-362349
; Sequence 362349, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
```

```
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; PRIOR FILING DATE: 2005-03-18
; PRIOR APPLICATION NUMBER: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 362349
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-362349
```

```
Query Match 66.0%; Score 13.2; DB 9; Length 19;
Best Local Similarity 61.1%; Pred. No. 8.6e+02;
Matches 11; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 2 AAAGTACCTTCACGCTT 19
    |||||:||||:|:|:|
Db 1 AAACUGACCUUACUACU 18
```

```
RESULT 17
US-11-083-784-1212199
; Sequence 1212199, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 1212199
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-1212199
```

```
Query Match 66.0%; Score 13.2; DB 9; Length 19;
Best Local Similarity 61.1%; Pred. No. 8.6e+02;
Matches 11; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 3 AACTGACCTTCACGCTT 20
    |||||:||||:|:|:|
Db 2 AACUGAUCUCCACUCCU 19
```

```
RESULT 18
US-11-083-784-1502436
; Sequence 1502436, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
```

```
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR APPLICATION NUMBER: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 1502436
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-1502436
```

```
Query Match 66.0%; Score 13.2; DB 9; Length 19;
Best Local Similarity 61.1%; Pred. No. 8.6e+02;
Matches 11; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 3 AACTGACCTTCACGCTT 20
    |||||:||||:|:|:|
Db 2 AACAGAGCUUCACACCU 19
```

```
RESULT 19
US-11-101-244-25073
; Sequence 25073, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 25073
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-25073
```

```
Query Match 64.0%; Score 12.8; DB 8; Length 19;
Best Local Similarity 68.8%; Pred. No. 1.3e+03;
Matches 11; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 GAACTGACCTTCACG 16
    |||||:||||:|:|:|
Db 4 GAAACUGCCCUCAUG 19
```

```
RESULT 20
US-11-101-244-25117
; Sequence 25117, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
```

; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/101,244  
; CURRENT FILING DATE: 2005-04-07  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 25117  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-101-244-25117

Query Match 64.0%; Score 12.8; DB 8; Length 19;  
Best Local Similarity 68.8%; Pred. No. 1.3e+03;  
Matches 11; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAACTGACCTTCACG 16  
|||||:|::||  
Db 3 GAACTGACCTTCACG 18

RESULT 21  
US-11-101-244-270762/c  
; Sequence 270762, Application US/11101244  
; Publication No. US20050246794A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmoon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/101,244  
; CURRENT FILING DATE: 2005-04-07  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 270762  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-101-244-270762

Query Match 64.0%; Score 12.8; DB 8; Length 19;  
Best Local Similarity 87.5%; Pred. No. 1.3e+03;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAACGTACCTTCACG 17  
|||||:|::||  
Db 16 AAACGTACCTTCACG 1

RESULT 22  
US-11-101-244-270835/c  
; Sequence 270835, Application US/11101244  
; Publication No. US20050246794A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmoon, Inc.  
; APPLICANT: Khvorova, Anastasia

; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/101,244  
; CURRENT FILING DATE: 2005-04-07  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 270835  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-101-244-270835

Query Match 64.0%; Score 12.8; DB 8; Length 19;  
Best Local Similarity 87.5%; Pred. No. 1.3e+03;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAACGTACCTTCACG 17  
|||||:|::||  
Db 16 AAACGTACCTTCACG 1

RESULT 23  
US-11-101-244-270911/c  
; Sequence 270911, Application US/11101244  
; Publication No. US20050246794A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmoon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/101,244  
; CURRENT FILING DATE: 2005-04-07  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 270911  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-101-244-270911

Query Match 64.0%; Score 12.8; DB 8; Length 19;  
Best Local Similarity 87.5%; Pred. No. 1.3e+03;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAACGTACCTTCACG 17  
|||||:|::||  
Db 16 AAACGTACCTTCACG 1

RESULT 24  
US-11-101-244-270975/c  
; Sequence 270975, Application US/11101244  
; Publication No. US20050246794A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmoon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela

```
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 134990US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 270975
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-270975
```

```
Query Match          64.0%; Score 12.8; DB 8; Length 19;
Best Local Similarity 87.5%; Pred. No. 1.3e+03;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAACGACCTTCACGC 17
   |||||
Db 19 AAACGACCTTCGGC 4
```

```
RESULT 25
US-11-101-244-466224
; Sequence 466224, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 134990US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 466224
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-466224
```

```
Query Match          64.0%; Score 12.8; DB 8; Length 19;
Best Local Similarity 62.5%; Pred. No. 1.3e+03;
Matches 10; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 5 CTGACCTTCACGCCTT 20
   |||||
Db 4 CUGACCGACCUCCUU 19
```

```
RESULT 26
US-11-101-244-940667
; Sequence 940667, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
```

```
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 134990US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 940667
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-940667
```

```
Query Match          64.0%; Score 12.8; DB 8; Length 19;
Best Local Similarity 56.2%; Pred. No. 1.3e+03;
Matches 9; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 5 CTGACCTTCACGCCTT 20
   |||||
Db 4 CUGCCUUCACACCUU 19
```

```
RESULT 27
US-11-101-244-1114497
; Sequence 1114497, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 134990US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 1114497
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-1114497
```

```
Query Match          64.0%; Score 12.8; DB 8; Length 19;
Best Local Similarity 75.0%; Pred. No. 1.3e+03;
Matches 12; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACG 16
   |||||
Db 1 GAAACCGACCUCAAG 16
```

```
RESULT 28
US-11-083-784-25073
; Sequence 25073, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
```



; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/083,784  
; CURRENT FILING DATE: 2005-03-18  
; PRIOR APPLICATION NUMBER: US/10/714,333  
; PRIOR FILING DATE: 2003-11-14  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 25073  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-083-784-25073

Query Match 64.0%; Score 12.8; DB 9; Length 19;  
Best Local Similarity 68.8%; Pred. No. 1.3e+03;  
Matches 11; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACG 16  
Db 4 GAAACUGCCCUCAUG 19  
|||||:| ||:| |

RESULT 29  
US-11-083-784-25117  
; Sequence 25117, Application US/11083784  
; Publication No. US20050245475A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmacon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/083,784  
; CURRENT FILING DATE: 2005-03-18  
; PRIOR APPLICATION NUMBER: US/10/714,333  
; PRIOR FILING DATE: 2003-11-14  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 25117  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-083-784-25117

Query Match 64.0%; Score 12.8; DB 9; Length 19;  
Best Local Similarity 68.8%; Pred. No. 1.3e+03;  
Matches 11; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAAACTGACCTTCACG 16  
Db 3 GAAACUGCCCUCAUG 18  
|||||:| ||:| |

RESULT 30  
US-11-083-784-270762/c  
; Sequence 270762, Application US/11083784  
; Publication No. US20050245475A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmacon, Inc.  
; APPLICANT: Khvorova, Anastasia

; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/083,784  
; CURRENT FILING DATE: 2005-03-18  
; PRIOR APPLICATION NUMBER: US/10/714,333  
; PRIOR FILING DATE: 2003-11-14  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 270762  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-083-784-270762

Query Match 64.0%; Score 12.8; DB 9; Length 19;  
Best Local Similarity 87.5%; Pred. No. 1.3e+03;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAACGTGACCTTCACGC 17  
Db 16 AAACGTGACCTTCGCG 1  
||||| ||||| ||

RESULT 31  
US-11-083-784-270835/c  
; Sequence 270835, Application US/11083784  
; Publication No. US20050245475A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmacon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/083,784  
; CURRENT FILING DATE: 2005-03-18  
; PRIOR APPLICATION NUMBER: US/10/714,333  
; PRIOR FILING DATE: 2003-11-14  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 270835  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-083-784-270835

Query Match 64.0%; Score 12.8; DB 9; Length 19;  
Best Local Similarity 87.5%; Pred. No. 1.3e+03;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAACGTGACCTTCACGC 17  
Db 16 AAACGTGACCTTCGCG 1  
||||| ||||| ||

RESULT 32  
US-11-083-784-270911/c  
; Sequence 270911, Application US/11083784  
; Publication No. US20050245475A1

```
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR FILING DATE: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 270911
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-270911
```

```
Query Match 64.0%; Score 12.8; DB 9; Length 19;
Best Local Similarity 87.5%; Pred. No. 1.3e+03;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 2 AAACGTGACCTTCACGC 17
    |||||
Db 16 AAACGTGACCTTCTGGC 1
```

## RESULT 33

```
US-11-083-784-270975/c
; Sequence 270975, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR FILING DATE: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 270975
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-270975
```

```
Query Match 64.0%; Score 12.8; DB 9; Length 19;
Best Local Similarity 87.5%; Pred. No. 1.3e+03;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 2 AAACGTGACCTTCACGC 17
    |||||
Db 19 AAACGTGACCTTCTGGC 4
```

## RESULT 34

```
US-11-083-784-466224
; Sequence 466224, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR FILING DATE: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 466224
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-466224
```

```
Query Match 64.0%; Score 12.8; DB 9; Length 19;
Best Local Similarity 62.5%; Pred. No. 1.3e+03;
Matches 10; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 5 CTGACCTTCACGCCTT 20
    |::|||::|||::|
Db 4 CUGACCUGCACCUU 19
```

## RESULT 35

```
US-11-083-784-940667
; Sequence 940667, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR FILING DATE: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 940667
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-940667
```

```
Query Match 64.0%; Score 12.8; DB 9; Length 19;
Best Local Similarity 56.2%; Pred. No. 1.3e+03;
Matches 9; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 5 CTGACCTTCACGCCTT 20
    |::|||::|||::|
Db 4 CUGCCCUUCACACCUU 19
```

Db 1 GAAACUGGCCUUCAACU 19

```

RESULT 38
US-11-101-244-355617
; Sequence 355617, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 355617
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-355617

```

```
Query Match      63.0%; Score 12.6; DB 8; Length 19;
Best Local Similarity 63.2%; Pred. No. 1.5e+03;
Matches 12; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
```

Qy 1 GAACTGACCTTCACGCCT 19  
||||| :  
Db 1 GAAACGGCCUUCAGACU 19

```

RESULT 39
US-11-101-244-452754
; Sequence 452754, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 452754
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-452754

```

Query Match 63.0%; Score 12.6; DB 8; Length 19;  
Best Local Similarity 52.6%; Pred. No. 1.5e+03;  
Matches 10: Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Qy 2 A A A C T G A C C T T C A C G C C T T 20  
| | | | | : | | : |  
pb 1 A U A C T A A C C T T C A U G C T T T 19

```

RESULT 36
US-11-083-784-1114497
/ Sequence 1114497, Application US/11083784
/ Publication No. US20050245475A1
/ GENERAL INFORMATION:
/ APPLICANT: Dharmacon, Inc.
/ APPLICANT: Khvorova, Anastasia
/ APPLICANT: Reynolds, Angela
/ APPLICANT: Leake, Devin
/ APPLICANT: Marshall, William
/ APPLICANT: Scaringe, Stephen
/ TITLE OF INVENTION: Functional and Hyperfunctional sirRNA
/ FILE REFERENCE: 13499US
/ CURRENT APPLICATION NUMBER: US/11/083,784
/ CURRENT FILING DATE: 2005-03-18
/ PRIOR APPLICATION NUMBER: US/10/714,333
/ PRIOR FILING DATE: 2003-11-14
/ PRIOR APPLICATION NUMBER: 60/502,050
/ PRIOR FILING DATE: 2003-09-10
/ PRIOR APPLICATION NUMBER: 60/426,137
/ PRIOR FILING DATE: 2002-11-14
/ NUMBER OF SEQ ID NOS: 1591911
/ SOFTWARE: Proprietary
/ SEQ ID NO 1114497
/ LENGTH: 19
/ TYPE: RNA
/ ORGANISM: Homo sapiens
US-11-083-784-1114497

```

Query Match	64.0%	Score 12.8;	DB 9;	Length 19;
Best Local Similarity	75.0%	Pred. No. 1.3e+03;		
Matches 12;	Conservative	2;	Mismatches 2;	Indels 0; Gaps 0;

Qy 1 GAAACTGACCTTCACG 16  
||| ||| ||| : |||  
Db 1 GAAACCGACCTTUCAG 16

```

RESULT 37
US-11-101-244-337258
; Sequence 337258, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 337258
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-337258

```

```
Query Match      63.0%; Score 12.6; DB 8; Length 19;
Best Local Similarity 57.9%; Pred. No. 1.5e+03;
Matches 11: Conservative 4; Mismatches 4; Indels 0; Gaps 0;
```

Qy 1 GAAACTGACCTTCACGCCT 19

```
RESULT 40
US-11-101-244-634465/c
; Sequence 634465, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 634465
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-634465
```

```
Query Match      63.0%; Score 12.6; DB 8; Length 19;
Best Local Similarity 78.9%; Pred. No. 1.5e+03;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY      2 AAACGTGACCTTCACGCCTT 20
        ||||| ||| ||| |||
Db      19 AAACGTGCCATCTCACCTT 1
```

Search completed: November 21, 2005, 05:22:41  
Job time : 280.506 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2005, 16:48:26 ; Search time 732.099 Seconds  
(without alignments)  
1552.890 Million cell updates/sec

Title: US-10-627-757-22

Perfect score: 20  
Sequence: 1 gagccaaacaggaaacaaac 20

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*

1: gb\_ba:\*

2: gb\_in:\*

3: gb\_env:\*

4: gb\_om:\*

5: gb\_ov:\*

6: gb\_pat:\*

7: gb\_ph:\*

8: gb\_pr:\*

9: gb\_ro:\*

10: gb\_sta:\*

11: gb\_sv:\*

12: gb\_un:\*

13: gb\_vi:\*

14: gb\_hg:\*

15: gb\_pl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	20	6	CQ7711174
C 2	20	100.0	1074	6	CQ7711157
C 3	20	100.0	4015	8	AF283519S5
C 4	20	100.0	196203	8	AL355355
C 5	20	100.0	208202	14	AC013446
C 6	19	95.0	128638	8	AC015968
C 7	19	95.0	181835	8	AC146227
C 8	19	95.0	182785	8	AC145937
C 9	18.4	92.0	116660	8	AC002074
C 10	18.4	92.0	137488	14	AC146426
C 11	18.4	92.0	163066	14	AC103900
C 12	18.4	92.0	164685	14	AC096680
C 13	18.4	92.0	185753	14	AC148263
C 14	18.4	92.0	188678	14	AC157650
C 15	18.4	92.0	192548	14	AC121536
C 16	18.4	92.0	215308	9	AC154455
C 17	18	90.0	195311	9	AC121585
C 18	18	90.0	207588	14	AC087162

C 19	18	90.0	211521	9	AC138101
C 20	18	90.0	240821	9	AC013622
C 21	17.4	87.0	38715	8	AC004237
C 22	17.4	87.0	59731	8	AC068442
C 23	17.4	87.0	60126	15	AP006630
C 24	17.4	87.0	72732	6	AX695578
C 25	17.4	87.0	143065	14	AC105375
C 26	17.4	87.0	146936	14	AP005010
C 27	17.4	87.0	148090	9	AC100737
C 28	17.4	87.0	148362	8	AC106811
C 29	17.4	87.0	152861	4	AC117937
C 30	17.4	87.0	158142	5	CR450814
C 31	17.4	87.0	159863	8	AL136218
C 32	17.4	87.0	179947	14	AC009786
C 33	17.4	87.0	180180	14	CR925835
C 34	17.4	87.0	184886	8	AC005358
C 35	17.4	87.0	189104	14	AC161229
C 36	17.4	87.0	189984	14	CR847954
C 37	17.4	87.0	196787	14	AC141646
C 38	17.4	87.0	200904	5	CR376854
C 39	17.4	87.0	204615	14	AC155199
C 40	17.4	87.0	219316	14	AC111278
C 41	17.4	87.0	221190	14	AC165077
C 42	17.4	87.0	245828	14	AC131349
C 43	17.4	87.0	248760	14	AC137456
C 44	17.4	87.0	250578	14	AC094142
C 45	17.4	87.0	251830	14	AC095714

ALIGNMENTS

RESULT 1					
CQ7711174					
LOCUS	CQ7711174	20 bp	DNA	linear	PAT 04-MAR-2004
DEFINITION	Sequence 22 from Patent EP1388590.				
ACCESSION	CQ7711174				
VERSION	CQ7711174.1	GI:45125307			
KEYWORDS					
SOURCE	synthetic construct				
ORGANISM	other sequences; artificial sequences.				
REFERENCE	1				
AUTHORS	Kouchi,Y., Masago,A. and Takahata,T.				
TITLE	Gene assay method for predicting glaucoma onset risk				
JOURNAL	Patent: EP 1388590-A 22 11-FEB-2004;				
FEATURES	Synsex Corporation (JP)				
source	location/Qualifiers				
	1..20				
	/organism="synthetic construct"				
	/mol_type="unassigned DNA"				
	/db_xref="taxon:33630"				

ORIGIN

Query Match	100.0%;	Score 20;	DB 6;	Length 20;
Best Local Similarity	100.0%;	Pred. No. 1.1e+02;		
Matches	20;	Conservative	0;	Mismatches 0;
		Indels	0;	Gaps 0;
Qy	1	GAGCCAAACGACCAAC	20	
Db	1	GAGCCAAACGACCAAC	20	
RESULT 2				
CQ771157/c				
LOCUS	CQ771157	1074 bp	DNA	linear
DEFINITION	Sequence 5 from Patent EP1388590.			
ACCESSION	CQ771157			
VERSION	CQ771157.1	GI:45125290		
KEYWORDS				
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;			

```

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1
REFERENCE
AUTHORS Kouchi,Y., Maaago,A. and Takahata,T.
TITLE Gene assay method for predicting glaucoma onset risk
JOURNAL Patent: EP 1388590-A 5 11-FEB-2004;
SYNOPSIS Synnex Corporation (JP)
FEATURES
source
1. .1074
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 100.0%; Score 20; DB 6; Length 1074;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGACCAAC 20
Db 784 GAGCCAAACAGGACCAAC 765

RESULT 3
AF283519S5/c
LOCUS AF283519S5 4015 bp DNA linear PRI 17-AUG-2000
DEFINITION Homo sapiens FIP2 gene, exons 7 and 8.
ACCESSION AF283523
VERSION AF283523.1 GI:9837251
KEYWORDS
SEGMENT 5 of 9
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 4015)
Li,D. and Roberts,R.
Human FIP-2: Genomic structure and mutational analysis in ARVD
patients
REFERENCE
AUTHORS Li,D. and Roberts,R.
TITLE Direct Submission
JOURNAL Submitted (30-JUN-2000) Medicine, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
FEATURES
source
1. .4015
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
exon 853. .926
exon 3477. .3629
/number=8

ORIGIN
Query Match 100.0%; Score 20; DB 8; Length 4015;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGACCAAC 20
Db 1136 GAGCCAAACAGGACCAAC 1117

RESULT 4
AL355355/c
LOCUS AL355355 196203 bp DNA linear PRI 18-MAY-2005
DEFINITION Human DNA sequence from clone RP11-730A19 on chromosome 10 Contains
the 5' end of a novel gene (DKFZP761F241) (FLJ20925 FLJ38473), a
ribosomal protein L5 (RPL5) pseudogene, the OPTN gene for

```

```

optineurin, a small nuclear ribonucleoprotein polypeptide G (SNRPG)
pseudogene, a pseudogene similar to part of COX10 homolog,
cytochrome c oxidase assembly protein heme A: farnesyltransferase
(yeast) (COX10), a novel pseudogene (FLJ10648 KIAA1525), a
ribosomal protein L36A (RPL36A) pseudogene, the 5' end of the
MCM10 gene for MCM10 minichromosome maintenance deficient 10 (S.
cerevisiae), a pseudogene similar to part of chromodomain protein Y
chromosome 1 (CDY1) and two CpG islands, complete sequence.
AL355355
AL355355.25 GI:16972859
HTG; CDY1; chromodomain; COX10; CpG island; DKFZP761F241; FLJ10648;
FLJ20925; FLJ38473; KIAA1525; MCM10; optineurin; OPTN; RPL36A;
RPL5; SNRPG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 196203)
Almeida,J.
Direct Submission
TITLE Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
JOURNAL Clone requests: clonerequest@sanger.ac.uk
COMMENT On Nov 16, 2001 this sequence version replaced gi:16214585.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep/
This sequence
was generated from part of bacterial clone contigs of human
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr10
RP11-730A19 is from the library RPCI-11.3 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: vegas@sanger.ac.uk
-----
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one subclone; and the assembly was confirmed by restriction digest,
except on the rare occasion of the clone being a YAC.
FEATURES
location/Qualifiers
1. .196203
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-730A19"
/clone_lib="RPCI-11.3"
misc_feature 1
/note="Clone left end: RP11-730A19"
gene join(complement(123034..123180),
complement(122612..122837),complement(122053..122130),
complement(97396..97628),complement(73202..73469),
complement(21866..22040),
complement(AL353586.14:43768..45820))
/locus_tag="RP11-347122.1"
/mRNA join(complement(123034..123180),
complement(122612..122837),complement(122053..122130),
complement(97396..97628),complement(73202..73469),
complement(21866..22040),
complement(AL353586.14:43768..45820))

```



```

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 208202)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 208202)
Waterston,R.H.
Direct Submission
Submitted (11-NOV-1999) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
On May 18, 2000 this sequence version replaced gi:6850545.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0513P21
----- Summary Statistics -----
Sequencing vector: M13; 84%
Chemistry: Dye-primer ET; 84% of reads
Assembly: Dye-terminator Big Dye; 16% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 196339 bases at least Q40
Consensus quality: 199886 bases at least Q30
Consensus quality: 202038 bases at least Q20
Insert size: 215000; agarose-fp
Insert size: 206002; sum-of-contigs
Quality coverage: 4.38 in Q20 bases; agarose-fp
Quality coverage: 4.26 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 23 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1
1699: contig of 1699 bp in length
1700: contig of unknown length
1800: contig of 1441 bp in length
3240: contig of unknown length
3241: contig of 2875 bp in length
6215: contig of 2740 bp in length
6316: contig of unknown length
9056: contig of 2689 bp in length
9156: contig of unknown length
11845: contig of 3647 bp in length
11945: contig of unknown length
15591: contig of unknown length
15692: contig of 2566 bp in length
18257: contig of unknown length
18358: contig of 5081 bp in length
23439: contig of unknown length
23539: contig of 4015 bp in length
27654: contig of unknown length
32488: contig of 4835 bp in length
32489: contig of 7089 bp in length
32589: contig of 3817 bp in length
36406: contig of unknown length
36505: contig of 4088 bp in length
40593: contig of unknown length
40594: contig of 7089 bp in length
47782: contig of unknown length
47883: contig of 7302 bp in length
55184: contig of unknown length
55284: contig of 8044 bp in length
55285: contig of unknown length
63329: contig of unknown length
63429: contig of 10673 bp in length
74101: contig of unknown length
74201: contig of unknown length
85050: contig of 10849 bp in length
74202

* 85051 85150: gap of unknown length
* 85151 100569: contig of 15419 bp in length
* 100570 100669: gap of unknown length
* 100670 116236: contig of 15567 bp in length
* 116237 116336: gap of unknown length
* 116337 131648: contig of 15312 bp in length
* 131649 131748: gap of unknown length
* 131749 148826: contig of 17078 bp in length
* 148827 148926: gap of unknown length
* 148927 175149: contig of 28223 bp in length
* 175150 175249: gap of unknown length
* 175250 208202: contig of 32953 bp in length.

FEATURES
Location/Qualifiers
source
1..208202
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-513P21"
misc_feature
1..1699
/note="assembly_name:Contig10"
gap
1700..1799
/estimated_length=unknown
misc_feature
1800..3240
/note="assembly_name:Contig11"
gap
3241..3340
/estimated_length=unknown
misc_feature
3341..6215
/note="assembly_name:Contig12"
gap
6216..6315
/estimated_length=unknown
misc_feature
6316..9055
/note="assembly_name:Contig13"
gap
9056..9155
/estimated_length=unknown
misc_feature
9156..11844
/note="assembly_name:Contig14"
gap
11845..11944
/estimated_length=unknown
misc_feature
11945..15591
/note="assembly_name:Contig15"
gap
15592..15691
/estimated_length=unknown
misc_feature
15692..18257
/note="assembly_name:Contig16"
gap
18258..18357
/estimated_length=unknown
misc_feature
18358..23438
/note="assembly_name:Contig17"
gap
23439..23538
/estimated_length=unknown
misc_feature
23539..27553
/note="assembly_name:Contig18"
gap
27554..27653
/estimated_length=unknown
misc_feature
27654..32488
/note="assembly_name:Contig19"
gap
32489..32588
/estimated_length=unknown
misc_feature
32589..36405
/note="assembly_name:Contig20"
gap
36406..36505
/estimated_length=unknown
misc_feature
36506..40593
/note="assembly_name:Contig21"
clone_end:T7
vector_side:right
40594..40693
/estimated_length=unknown
gap
40694..47782
/note="assembly_name:Contig22"
47783..47882
/estimated_length=unknown

```





```
repeat_region /rpt_family="(TATG)n"
repeat_region 402..663
repeat_region /rpt_family="Alu"
repeat_region 991..1295
repeat_region /rpt_family="Alu"
repeat_region 2025..2055
repeat_region /rpt_family="(TA)n"
repeat_region 2057..2088
repeat_region /rpt_family="(GA)n"
repeat_region 2089..2363
repeat_region /rpt_family="Alu"
repeat_region 2441..2542
repeat_region /rpt_family="MIR"
repeat_region 2569..3173
repeat_region /rpt_family="ERV1"
repeat_region 5618..5926
repeat_region /rpt_family="Alu"
repeat_region 5967..6231
repeat_region /rpt_family="Alu"
repeat_region 6418..6949
repeat_region /rpt_family="MER1_type"
repeat_region 7921..8227
repeat_region /rpt_family="Alu"
repeat_region 8398..8478
repeat_region /rpt_family="MIR"
repeat_region 8558..8696
repeat_region /rpt_family="MIR"
repeat_region 11213..11247
repeat_region /rpt_family="(A)n"
repeat_region 14788..14814
repeat_region /rpt_family="(TTA)n"
repeat_region 14815..15576
repeat_region /rpt_family="L1"
repeat_region 15574..16532
repeat_region /rpt_family="L1"
repeat_region 16762..17064
repeat_region /rpt_family="Alu"
repeat_region 20843..22355
repeat_region /rpt_family="L1"
repeat_region 22359..22696
repeat_region /rpt_family="L1"
repeat_region 22706..23010
repeat_region /rpt_family="Alu"
repeat_region 23834..24127
repeat_region /rpt_family="Alu"
repeat_region 24154..24601
repeat_region /rpt_family="L1"
repeat_region 25597..26090
repeat_region /rpt_family="L1"
repeat_region 26229..26249
repeat_region /rpt_family="AT_rich"
repeat_region 29072..29257
repeat_region /rpt_family="L1"
repeat_region 29450..29851
repeat_region /rpt_family="L1"
repeat_region 29841..30049
repeat_region /rpt_family="L1"
repeat_region 30050..30121
repeat_region /rpt_family="Alu"
repeat_region 30122..30693
repeat_region /rpt_family="L1"
repeat_region 30694..30996
repeat_region /rpt_family="Alu"
repeat_region 30997..31396
repeat_region /rpt_family="L1"
repeat_region 31397..31700
repeat_region /rpt_family="Alu"
repeat_region 31701..32174
repeat_region /rpt_family="L1"
repeat_region 32198..32500
repeat_region /rpt_family="Alu"
repeat_region 32502..32924
repeat_region /rpt_family="MaLR"
```

```
repeat_region 32939..32994
repeat_region /rpt_family="A-rich"
repeat_region 34831..35126
repeat_region /rpt_family="Alu"
repeat_region 35193..35649
repeat_region /rpt_family="L1"
repeat_region 35863..36526
repeat_region /rpt_family="ERV1"
repeat_region 36527..36951
repeat_region /rpt_family="ERV1"
repeat_region 37518..37562
repeat_region /rpt_family="AT_rich"
repeat_region 37630..37897
repeat_region /rpt_family="L1"
repeat_region 38642..38967
repeat_region /rpt_family="ERV1"
repeat_region 39307..39606
repeat_region /rpt_family="Alu"
repeat_region 40032..40362
repeat_region /rpt_family="MaLR"
repeat_region 40987..41230
repeat_region /rpt_family="L1"
repeat_region 41658..41954
repeat_region /rpt_family="Alu"
repeat_region 42028..42085
repeat_region /rpt_family="L1"
repeat_region 42086..42361
repeat_region /rpt_family="L1"
repeat_region 42465..43084
repeat_region /rpt_family="L1"
repeat_region 43084..44825
repeat_region /rpt_family="L1"
misc_feature 43512..43698
/note="unresolved homopolymeric repeat"
```

Query Match 95.0%; Score 19; DB 8; Length 128638;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGACCAAA 19  
|||||  
Db 64619 GAGCCAAACAGGACCAAA 64601

## RESULT 7

AC146227/c AC146227 181835 bp DNA linear PRI 29-OCT-2003  
LOCUS Pan troglodytes BAC clone RP43-32J19 from 7, complete sequence.  
DEFINITION AC146227  
ACCESSION AC146227  
VERSION AC146227.2 GI:37951460  
KEYWORDS HTG.  
SOURCE Pan troglodytes (chimpanzee)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Pan.  
REFERENCE 1 (bases 1 to 181835)  
AUTHORS Hodges,J., Haakenson,W. and Bielicki,L.  
TITLE The sequence of Pan troglodytes BAC clone RP43-32J19  
JOURNAL Unpublished (2001)  
REFERENCE 2 (bases 1 to 181835)  
AUTHORS Sulston,J.E. and Wilson,R.  
TITLE Sequencing of Pan troglodytes  
JOURNAL Unpublished (2001)  
REFERENCE 3 (bases 1 to 181835)  
AUTHORS Wilson,R.K.  
TITLE Direct Submission  
JOURNAL Submitted (01-AUG-2003) Genetics, Genome Sequencing Center, 4444  
Forest Park Parkway, St. Louis, MO 63108, USA  
REFERENCE 4 (bases 1 to 181835)  
AUTHORS Wilson,R.K.  
TITLE Direct Submission  
JOURNAL Submitted (23-AUG-2003) Genetics, Genome Sequencing Center, 4444

```

REFERENCE
AUTHORS      Forest Park Parkway, St. Louis, MO 63108, USA
TITLE        5 (bases 1 to 181835)
JOURNAL      Wilson,R.K.
SOURCE       Direct Submission
COMMENT      Submitted (24-OCT-2003) Genetics, Genome Sequencing Center, 4444
              Forest Park Parkway, St. Louis, MO 63108, USA
              6 (bases 1 to 181835)
              Wilson,R.
              Direct Submission
              Submitted (29-OCT-2003) Department of Genetics, Washington
              University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
              On Oct 24, 2003 this sequence version replaced gi:33387180.
              ----- Genome Center
              Center: Washington University Genome Sequencing Center
              Center code: WUGSC
              Web site: http://genome.wustl.edu
              Contact: submissions@watson.wustl.edu
              ----- Summary Statistics
              -----
              -----
              NOTICE: This sequence may not represent the entire insert of this
              clone. It may be shorter because we only sequence overlapping
              clone sections once, or longer because we provide a small overlap
              between neighboring data submissions.

              This sequence was finished as follows unless otherwise noted:
              all regions were double stranded, sequenced with an alternate
              chemistry, or covered by high quality data (i.e., phred quality >=
              30); an attempt was made to resolve all sequencing problems, such
              as compressions and repeats; all regions were covered by sequence
              from more than one subclone; and the assembly was confirmed by
              restriction digest.

              MAPPING INFORMATION:
              Mapping information for this clone was provided by Dr. Wes Warren,
              Department of Genetics, Washington University, St. Louis MO. For
              additional information about the map position of this sequence, see
              http://genome.wustl.edu

              SOURCE INFORMATION:
              The RPCI-43 BAC Library has been constructed by Chung-Li Shu. DNA
              was isolated from white blood cells obtained from a male chimpanzee
              (Pan troglodytes, 'Clint', Yerkes #C0471; birthdate: 6-6-80). The
              clone and detailed information can be obtained from ResGen
              (http://www.resgen.com) or Pieter de Jong and co-workers at
              http://www.bacpac.chori.org.

              NEIGHBORING SEQUENCE INFORMATION:
              This sequence is the entire insert of the clone.
              Location/Qualifiers
                1. .181835
                  /organism="Pan troglodytes"
                  /mol_type="genomic DNA"
                  /db_xref="taxon:9598"
                  /chromosome="7"
                  /map="7"
                  /clone="RP43-32J19"
                  /clone_lib="RPCI-43"
                30903..30948
                  /note="Sequence derived from one plasmid subclone."
                64482..64541
                  /note="Sequence derived from one plasmid subclone."
                76618..77020
                  /note="Unresolved tandem repeat."
                87893..88110
                  /note="Unresolved simple sequence repeat."

              ORIGIN
              Query Match          95.0%; Score 19; DB 8; Length 181835;
              Best Local Similarity 100.0%; Pred. No. 1.1e-02;
              Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

              QY      1 GAGCCAAACAGGAACCAA 19
              Db      19279 GAGCCAAACAGGAACCAA 19261
              RESULT 8
              AC145937      182785 bp      DNA      linear      PRI 27-APR-2005
              AC145937      Pan troglodytes BAC clone RP43-10P16 from chromosome 7, complete
              AC145937      sequence.
              AC145937      AC145937
              AC145937.2      GI:36016669
              HTG.
              SOURCE       Pan troglodytes (chimpanzee)
              ORGANISM     Pan troglodytes
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
              Homnidae; Pan.
              1 (bases 1 to 182785)
              Tomlinson,C., Cotton,M. and Kozlowicz,A.
              The sequence of Pan troglodytes BAC clone RP43-10P16
              Unpublished (2001)
              2 (bases 1 to 182785)
              Wilson,R.K.
              Direct Submission
              Submitted (01-AUG-2003) Genetics, Genome Sequencing Center, 4444
              Forest Park Parkway, St. Louis, MO 63108, USA
              3 (bases 1 to 182785)
              Wilson,R.
              Direct Submission
              Submitted (26-SEP-2003) Department of Genetics, Washington
              University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
              4 (bases 1 to 182785)
              Wilson,R.K.
              Direct Submission
              Submitted (27-APR-2005) Washington University School of Medicine,
              Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO
              63108, USA
              ----- Genome Center
              Center: Washington University Genome Sequencing Center
              Center code: WUGSC
              Web site: http://genome.wustl.edu
              Contact: submissions@watson.wustl.edu
              ----- Summary Statistics
              -----
              Center project name: C_PT010P16

              NOTICE:
              This sequence was finished as follows unless otherwise noted:
              all regions were double stranded, sequenced with an alternate
              chemistry, or covered by high quality data (i.e., phred quality >=
              30); an attempt was made to resolve all sequencing problems, such
              as compressions and repeats; all regions were covered by sequence
              from more than one subclone; and the assembly was confirmed by
              restriction digest.

              MAPPING INFORMATION:
              Mapping information for this clone was provided by Dr. Wes Warren,
              Department of Genetics, Washington University, St. Louis MO. For
              additional information about the map position of this sequence, see
              http://genome.wustl.edu

              SOURCE INFORMATION:
              The RPCI-43 BAC Library has been constructed by Chung-Li Shu. DNA
              was isolated from white blood cells obtained from a male chimpanzee
              (Pan troglodytes, 'Clint', Yerkes #C0471; birthdate: 6-6-80). The
              clone and detailed information can be obtained from ResGen
              (http://www.resgen.com) or Pieter de Jong and co-workers at
              http://www.bacpac.chori.org.

              NEIGHBORING SEQUENCE INFORMATION:
              This sequence is the entire insert of the clone.
              Location/Qualifiers
                1. .181835
                  /organism="Pan troglodytes"
                  /mol_type="genomic DNA"
                  /db_xref="taxon:9598"
                  /chromosome="7"
                  /map="7"
                  /clone="RP43-32J19"
                  /clone_lib="RPCI-43"
                30903..30948
                  /note="Sequence derived from one plasmid subclone."
                64482..64541
                  /note="Sequence derived from one plasmid subclone."
                76618..77020
                  /note="Unresolved tandem repeat."
                87893..88110
                  /note="Unresolved simple sequence repeat."

              ORIGIN
              Query Match          95.0%; Score 19; DB 8; Length 181835;
              Best Local Similarity 100.0%; Pred. No. 1.1e-02;
              Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

clone and detailed information can be obtained from ResGen  
(http://www.resgen.com) or Pieter de Jong and co-workers at  
http://www.bacpac.chori.org.

This sequence is the entire insert of the clone.

```

FEATURES
    source
        1..182785
            Location/Qualifiers
                /organism="Pan troglodytes"
                /mol_type="genomic DNA"
                /db_xref="taxon:9598"
                /chromosome="7"
                /clone="RP43-10P16"
                /clone_lib="RPCI-43"
                14981..15169
                /note="Unresolved simple sequence repeat."
                26038..26707
                /note="Unresolved tandem repeat."

    unsure
    unsure

ORIGIN
    Query Match          95.0%; Score 19; DB 8; Length 182785;
    Best Local Similarity 100.0%; Pred. NO. 1.1e+02;
    Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GAGCCAAACAGGACCAAA 19
        |||||
Db      83458 GAGCCAAACAGGACCAAA 83476

RESULT 9
AC002074/c
LOCUS   AC002074          116660 bp      DNA      linear      PRI 25-FEB-2004
DEFINITION Homo sapiens BAC clone GS1-56H18 from 7, complete sequence.
ACCESSION AC002074
VERSION   AC002074.2 GI:18093043
KEYWORDS HTG.
SOURCE   Homo sapiens
    ORGANISM Homo sapiens (human)
        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
        Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
        Hominoidea; Homo.
REFERENCE 1 (bases 1 to 116660)
AUTHORS Hillier,L.M., Fulton,R.S., Fulton,L.A., Graves,T.A., Pepin,K.H.,
Wagner-McPherson,C., Layman,D., Maas,J., Jaeger,S., Walker,R.,
Wylie,K., Sekhon,M., Becker,M.C., O'Laughlin,M.D., Schaller,M.E.,
Fewell,G.A., Delehaunty,K.D., Miner,T.L., Nash,W.E., Cordes,M.,
Du,H., Sun,H., Edwards,J., Bradshaw-Cordum,H., Ali,J., Andrews,S.,
Isak,A., Vanbrunt,A., Nguyen,C., Du,F., Lamar,B., Courtney,L.,
Kalicki,J., Ozersky,P., Bielicki,L., Scott,K., Holmes,A.,
Harkins,R., Harris,A., Strong,C.M., Hou,S., Tomlinson,C.,
Dauphin-Kohlberg,S., Kozlowski-Reilly,A., Leonard,S., Rohlfing,T.,
Rock,S.M., Tin-Wollam,A.M., Abbott,A., Minx,P., Maupin,R.,
Strommatt,C., Latreille,P., Miller,N., Johnson,D., Murray,J.,
Woesner,J.P., Wendt,M.C., Yang,S.P., Schultz,B.R., Wallis,J.W.,
Spieth,J., Bieri,T.A., Nelson,J.O., Berkowicz,N., Wohldmann,P.E.,
Cook,L.L., Hickenbotham,M.T., Eldred,J., Williams,D., Bedell,J.A.,
Mardis,E.R., Clifton,S.W., Chissoe,S.L., Maizra,M.A., Raymond,C.,
Haugen,E., Gillett,W., Zhou,Y., James,R., Phelps,K., Radanoto,S.,
Bubb,K., Simms,E., Levy,R., Clendenning,J., Kaul,R., Kent,W.J.,
Furey,T.S., Baertsch,R.A., Brent,M.R., Keibler,E., Flince,P.,
Bork,P., Suyama,M., Bailey,J.A., Portnoy,M.E., Torrents,D.,
Chirwall,A.T., Gish,W.R., Eddy,S.R., McPherson,J.D., Olson,M.V.,
Eichler,E.E., Green,E.D., Waterston,R.H. and Wilson,R.K.
    The DNA sequence of human chromosome 7
    Nature 424 (6945), 157-164 (2003)
REFERENCE 2 (bases 1 to 116660)
AUTHORS Smith,A. and Wamsley,P.
TITLE   The sequence of Homo sapiens BAC clone GS1-56H18
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 116660)
AUTHORS Waterston,R.
TITLE   Direct Submission
JOURNAL Submitted (12-MAY-1997) Department of Genetics, Washington

```

# REFERENCE

## AUTHORS

## JOURNAL

## REFERENCE

## AUTHORS

## JOURNAL

## REFERENCE

## AUTHORS

## JOURNAL

## REFERENCE

## AUTHORS

## JOURNAL

## COMMENT

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
4 (bases 1 to 116660)  
Waterston,R.  
Direct Submission  
Submitted (03-FEB-2000) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
5 (bases 1 to 116660)  
Waterston,R.  
Direct Submission  
Submitted (04-FEB-2000) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
6 (bases 1 to 116660)  
Waterston,R.  
Direct Submission  
Submitted (09-JAN-2002) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
7 (bases 1 to 116660)  
Wilson,R.  
Direct Submission  
Submitted (30-JAN-2004) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Jan 9, 2002 this sequence version replaced gi:2078467.

Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu  
Contact: sapiens@watson.wustl.edu  
----- Summary Statistics  
-----  
Center project name: H\_GS056H18  
-----

NOTICE: This sequence may not represent the entire insert of this  
clone. It may be shorter because we only sequence overlapping  
clone sections once, or longer because we provide a small overlap  
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by sequence  
from more than one subclone; and the assembly was confirmed by  
restriction digest.

MAPPING INFORMATION:  
The sequence of this clone was established as part of a mapping and  
sequencing collaboration between the NHGRI Chromosome 7 Mapping  
Project (Eric D. Green, Director), John D. McPherson in the  
Department of Genetics (Washington University), and the Washington  
University Genome Sequencing Center. For additional information  
about the map position of this sequence, see  
http://www.nhgri.nih.gov/DIR/GRB/CHR7, send  
mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu

SOURCE INFORMATION:  
This clone is from the first BAC library from Genome Systems, Inc.  
(http://www.genomesystems.com).

Cell line: lymphoblastoid  
Haplotype: two  
VECTOR: pBelOBAC  
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the right is GS1-542D18. Actual start of  
this clone is at base position 1 of GS1-56H18 actual end is at base  
position 19031 of GS1-542D18.

Location/Qualifiers  
1..116660  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="7"  
/map="7"

FEATURES  
source

```

/clone="GS1-56H18"
/clone_lib="GSGAC1"
989. .1125
/rpt_family="L2"
repeat_region
1918. .2625
/rpt_family="L2"
repeat_region
4267. .4314
/rpt_family="CT-rich"
repeat_region
5690. .5756
/rpt_family="MIR"
repeat_region
6247. .6347
/rpt_family="L2"
repeat_region
7190. .7213
/rpt_family="(T)n"
repeat_region
7646. .7890
/rpt_family="L2"
repeat_region
8464. .8490
/rpt_family="(T)n"
repeat_region
9447. .9470
/rpt_family="(TC)n"
repeat_region
9471. .9563
/rpt_family="MIR"
repeat_region
9701. .9783
/rpt_family="L2"
repeat_region
9784. .9924
/rpt_family="MIR"
repeat_region
10025. .10245
/rpt_family="MIR"
repeat_region
12409. .12443
/rpt_family="(CA)n"
repeat_region
12552. .12589
/rpt_family="(TG)n"
repeat_region
12590. .12717
/rpt_family="GA-rich"
repeat_region
15013. .15041
/rpt_family="(GA)n"
repeat_region
15635. .16242
/rpt_family="L2"
repeat_region
16243. .16555
/rpt_family="Alu"
repeat_region
16556. .16729
/rpt_family="L2"
repeat_region
16798. .17094
/rpt_family="L2"
repeat_region
17156. .17209
/rpt_family="L2"
repeat_region
17212. .17379
/rpt_family="ERV1"
repeat_region
17627. .17745
/rpt_family="ERV1"
repeat_region
17746. .17901
/rpt_family="L2"
repeat_region
17902. .18211
/rpt_family="Alu"
repeat_region
18212. .19386
/rpt_family="L2"
repeat_region
19387. .19724
/rpt_family="L1"
repeat_region
19725. .20023
/rpt_family="Alu"
repeat_region
20024. .20267
/rpt_family="L1"
repeat_region
20268. .20370
/rpt_family="L2"
repeat_region
20894. .21286
/rpt_family="MaLR"
repeat_region
21438. .21532
/rpt_family="L2"
repeat_region
22078. .22158
/rpt_family="MER103"
repeat_region
22521. .22889
/rpt_family="MaLR"
repeat_region
23417. .23620

```

```

/rpt_family="MaLR"
24404. .24540
/rpt_family="L2"
repeat_region
26790. .26858
/rpt_family="MIR"
repeat_region
27129. .27510
/rpt_family="L2"
repeat_region
28374. .28422
/rpt_family="MIR"
repeat_region
28715. .29100
/rpt_family="L2"

Query Match      92.0%; Score 18.4; DB 8; Length 116660;
Beat Local Similarity 95.0%; Pred. NO. 2.2e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGACCAAC 20
Db 44172 GAACCAACAGGACCAAC 44153

RESULT 10
AC146426/c
LOCUS
DEFINITION
AC146426
AC146426.1 GI:33621002
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
Wilson,R.K.
TITLE
The sequence of Pan troglodytes clone
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 137488)
AUTHORS
Wilson,R.K.
TITLE
Direct Submission
JOURNAL
Submitted (13-AUG-2003) Genetics, Genome Sequencing Center, 4444
Forest Park Parkway, St. Louis, MO 63108, USA

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: C_PT025E06

* NOTE: This is a 'working draft' sequence. It currently
* consists of 43 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1415: contig of 1415 bp in length
* 1416 1515: gap of unknown length
* 1516 3074: contig of 1559 bp in length
* 3075 3174: gap of unknown length
* 3175 4688: contig of 1514 bp in length
* 4689 4788: gap of unknown length
* 4789 6089: contig of 1301 bp in length
* 6090 6190: gap of unknown length
* 6190 7205: contig of 1016 bp in length
* 7206 7306: gap of unknown length
* 7306 8660: contig of 1355 bp in length
* 8661 8761: gap of unknown length
* 8761 10761: contig of 2001 bp in length

```

\* 10762 10861: gap of unknown length  
\* 10862 12312: contig of 1651 bp in length  
\* 12513 12612: gap of unknown length  
\* 13796 13796: contig of 1184 bp in length  
\* 13797 13896: gap of unknown length  
\* 13897 15618: contig of 1722 bp in length  
\* 15619 15718: gap of unknown length  
\* 15719 16912: contig of 1194 bp in length  
\* 16913 17012: gap of unknown length  
\* 17013 19472: contig of 2460 bp in length  
\* 19473 19572: gap of unknown length  
\* 19573 20915: contig of 1343 bp in length  
\* 20916 21015: gap of unknown length  
\* 21016 22991: contig of 1976 bp in length  
\* 22992 23091: gap of unknown length  
\* 23092 25146: contig of 2055 bp in length  
\* 25147 25246: gap of unknown length  
\* 25247 28062: contig of 2816 bp in length  
\* 28063 28162: gap of unknown length  
\* 28163 30604: contig of 2442 bp in length  
\* 30605 30704: gap of unknown length  
\* 30705 33055: contig of 2351 bp in length  
\* 33056 33155: gap of unknown length  
\* 33156 34963: contig of 1808 bp in length  
\* 34964 35063: gap of unknown length  
\* 35064 37276: contig of 2213 bp in length  
\* 37277 37376: gap of unknown length  
\* 37377 40018: contig of 2642 bp in length  
\* 40019 40118: gap of unknown length  
\* 40119 41827: contig of 1709 bp in length  
\* 41828 41927: gap of unknown length  
\* 41928 44382: contig of 2455 bp in length  
\* 44383 44482: gap of unknown length  
\* 44483 47276: contig of 2794 bp in length  
\* 47277 47376: gap of unknown length  
\* 47377 49784: contig of 2408 bp in length  
\* 49785 49884: gap of unknown length  
\* 49885 53766: contig of 3882 bp in length  
\* 53767 53866: gap of unknown length  
\* 53867 57583: contig of 3717 bp in length  
\* 57584 57683: gap of unknown length  
\* 57684 61097: contig of 3414 bp in length  
\* 61098 61197: gap of unknown length  
\* 61198 64210: contig of 3013 bp in length  
\* 64211 64310: gap of unknown length  
\* 64311 67891: contig of 3581 bp in length  
\* 67892 67991: gap of unknown length  
\* 67992 72829: contig of 4838 bp in length  
\* 72830 72929: gap of unknown length  
\* 72930 76301: contig of 3372 bp in length  
\* 76302 76401: gap of unknown length  
\* 76402 79571: contig of 3170 bp in length  
\* 79572 79671: gap of unknown length  
\* 79672 84538: contig of 4867 bp in length  
\* 84539 84638: gap of unknown length  
\* 84639 88786: contig of 4148 bp in length  
\* 88787 88886: gap of unknown length  
\* 88887 94837: contig of 5951 bp in length  
\* 94838 94937: gap of unknown length  
\* 94938 102832: contig of 7955 bp in length  
\* 102893 102932: gap of unknown length  
\* 102993 111296: contig of 8304 bp in length  
\* 111297 111396: gap of unknown length  
\* 111397 117659: contig of 6273 bp in length  
\* 117660 117769: gap of unknown length  
\* 117770 125482: contig of 7713 bp in length  
\* 125483 125582: gap of unknown length  
\* 125583 134891: contig of 9309 bp in length  
\* 134892 134991: gap of unknown length  
\* 134992 136235: contig of 1244 bp in length  
\* 136236 136335: gap of unknown length  
\* 136336 137488: contig of 1153 bp in length.

Location/Qualifiers

FEATURES

source  
1. .137488  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
/chromosome="UNK"  
/clone="RP43-25B6"  
1. .1415  
/note="assembly\_name:Contig13"  
1416. .1515  
/estimated\_length=unknown  
1516. .3074  
/note="assembly\_name:Contig21"  
3075. .3174  
/estimated\_length=unknown  
3175. .4688  
/note="assembly\_name:Contig23"  
4689. .4788  
/estimated\_length=unknown  
4789. .6089  
/note="assembly\_name:Contig24"  
6090. .6189  
/estimated\_length=unknown  
6190. .7205  
/note="assembly\_name:Contig25"  
7206. .7305  
/estimated\_length=unknown  
7306. .8660  
/note="assembly\_name:Contig26"  
8661. .8760  
/estimated\_length=unknown  
8761. .10761  
/note="assembly\_name:Contig27"  
10762. .10861  
/estimated\_length=unknown  
10862. .12512  
/note="assembly\_name:Contig28"  
12513. .12612  
/estimated\_length=unknown  
12613. .13796  
/note="assembly\_name:Contig29"  
13797. .13896  
/estimated\_length=unknown  
13897. .15618  
/note="assembly\_name:Contig30"  
15619. .15718  
/estimated\_length=unknown  
15719. .16912  
/note="assembly\_name:Contig31"  
16913. .17012  
/estimated\_length=unknown  
17013. .19472  
/note="assembly\_name:Contig32"  
19473. .19572  
/estimated\_length=unknown  
19573. .20915  
/note="assembly\_name:Contig33"  
20916. .21015  
/estimated\_length=unknown  
21016. .22991  
/note="assembly\_name:Contig34"  
22992. .23091  
/estimated\_length=unknown  
23092. .25146  
/note="assembly\_name:Contig35"  
25147. .25246  
/estimated\_length=unknown  
25247. .28062  
/note="assembly\_name:Contig36"  
28063. .28162  
/estimated\_length=unknown  
28163. .30604  
/note="assembly\_name:Contig37"  
30605. .30704

misc\_feature

gap

misc\_feature

gap

misc\_feature

gap

misc\_feature

gap

misc\_feature

gap

misc\_feature

gap

misc\_feature

gap

misc\_feature

gap

misc\_feature

gap

misc\_feature

gap

misc\_feature

gap

misc\_feature

gap

misc\_feature

gap

misc\_feature

gap

misc\_feature

gap

misc\_feature

gap

misc\_feature

gap

```

misc_feature      /estimated length=unknown
gap              30705..33055
                /note="assembly_name:Contig38"
                33056..33155

Query Match      92.0%; Score 18.4; DB 14; Length 137488;
Best Local Similarity 95.0%; Pred. No. 2.2e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAAC 20
    ||| ||||| ||||| |||||
Db 18321 GAACCAACAGGAACCAAC 18302

RESULT 11
AC103900/c
LOCUS          163066 bp      DNA      linear      HTG 13-JUN-2002
DEFINITION    Pan troglodytes clone RP43-161C11, WORKING DRAFT SEQUENCE, 2
ordered pieces.
AC103900
VERSION       AC103900.2 GI:21397300
KEYWORDS      HTG; HTGS PHASE2; HTGS DRAFT.
SOURCE        Pan troglodytes (chimpanzee)
ORGANISM      Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Pan.
REFERENCE     1 (bases 1 to 163066)
AUTHORS      Akther, N., Antonellis, A., Ayele, K., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C.,
Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J.,
Haghighi, P., Hansen, N., Ho, S.-L., Idol, J.R., Karlins, E., Laric, P.,
Lee-Lin, S.-Q., Legaspi, R., Maduro, Q.L., Maduro, V.B.,
Margulies, E.H., Masello, C., Maskeri, B., Mastrian, S.D.,
McCloskey, J.C., McDowell, J., Paquirigan, C., Pearson, R.,
Portnoy, M.E., Prasad, A., Schueter, M.G., Stantropop, S., Thomas, J.W.,
Thomas, P.J., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,
Wetherby, K.B., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 163066)
Green, E.D.
Direct Submission
Submitted (13-JUN-2002) NIH Intramural Sequencing Center, 8717
Grovmont Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 163066)
Green, E.D.
Direct Submission
Submitted (13-JUN-2002) NIH Intramural Sequencing Center, 8717
Grovmont Circle, Gaithersburg, MD 20877, USA
On Jun 13, 2002 this sequence version replaced gi:17155052.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.zoo@nhgri.nih.gov
----- Project Information
Center project name: any
Center clone name: 161C11

The sequence data in this record represents an 'enhanced'
version of a Phase 2 submission. Specifically, the indicated
order and orientation of each sequence contig has been
established using one or more of the following: read-pair
data from individual subclones, overlaps with neighboring
clones, alignment with available reference sequence (e.g.,
human), and/or confirmation by PCR testing. In addition,
the sequence assembly is based on at least 8X average
coverage in Q20 bases and has been reviewed to rule out
gross misassemblies, the low-quality ends of sequence
contigs have been trimmed away, and each base is associated
with a Phrap-derived quality score.
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 162872 bases at least Q40
Consensus quality: 162949 bases at least Q30
Consensus quality: 162965 bases at least Q20
Insert size: 138000; agarose-fp
Insert size: 124000; pulse-field-gel
Quality coverage: 12.74x in Q20 bases; agarose-fp
Quality coverage: 14.17x in Q20 bases; pulse-field-gel
Quality coverage: 10.79x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 1 68998: contig of 68998 bp in length
* 68999 163066: contig of 93968 bp in length.
*
FEATURES             Location/Qualifiers
     source           1..163066
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="RP43-161C11"
                     /clone_lib="RP43"
     misc_feature     1..68998
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     gap              68999..69098
                     /estimated_length=unknown
                     69099..163066
                     /note="assembly_fragment
                     clone_end:177
                     vector_side:right"
     misc_feature     76422..163066
                     /note="clone overlaps with GenBank Accession Number
                     AC103900 clone RP43-161C11 (center project name anv)"
ORIGIN
Query Match      92.0%; Score 18.4; DB 14; Length 163066;
Best Local Similarity 95.0%; Pred. No. 2.1e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAAC 20
    ||| ||||| ||||| |||||
Db 145564 GAACCAACAGGAACCAAC 145545

RESULT 12
AC096680/c
LOCUS          164685 bp      DNA      linear      HTG 24-SEP-2003
DEFINITION    Pan troglodytes clone RP43-25B6, WORKING DRAFT SEQUENCE.
AC096680
ACCESSION     AC096680
VERSION       AC096680.3 GI:35062811
KEYWORDS      HTG; HTGS PHASE2; HTGS DRAFT.
SOURCE        Pan troglodytes (chimpanzee)
ORGANISM      Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Pan.
REFERENCE     1 (bases 1 to 164685)
AUTHORS      Antonellis, A., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S., Carriaga, K.,
Chu, G., Coleman, B., Coleman, H., Engle, J., Granite, S., Guan, X.,
Gupta, J., Haghighi, P., Han, J., Hansen, N., Ho, S.-L., Hu, P.,

```

Hurle,B., Idol,J.R., Karlins,E., Kwong,P., Laric,P., Lee-Lin,S.-Q.,  
Legaspi,R., Maduro,O.L., Maduro,V.B., Margulies,E.H., Masiello,C.,  
Maskeri,B., McDowell,J., Peguirigan,C., Pearson,R., Portnoy,M.E.,  
Prasad,A., Reddi-Dugue,N., Schandler,K., Schueler,M.G., Shah,K.,  
Sison,C., Stantripop,S., Thomas,J.W., Thomas,P.J., Tsipouri,V.,  
Vogt,J.L., Wetherby,K.D., Young,A. and Green,E.D.  
NISC Comparative Sequencing Initiative

# JOURNAL

## REFERENCE

### AUTHORS

### JOURNAL

### REFERENCE

### AUTHORS

### JOURNAL

### COMMENT

Unpublished  
2. (bases 1 to 164685)  
Green,E.D.  
Direct Submission  
Submitted (21-SEP-2001) NIH Intramural Sequencing Center, 8717  
Grovmont Circle, Gaithersburg, MD 20877, USA  
3. (bases 1 to 164685)  
Green,E.D.  
Direct Submission  
Submitted (24-SEP-2003) NIH Intramural Sequencing Center, 8717  
Grovmont Circle, Gaithersburg, MD 20877, USA  
On Sep 24, 2003 this sequence version replaced gi:21397303.  
----- Genome Center  
Center: NIH Intramural Sequencing Center  
Center code: NISC  
Web site: <http://www.nisc.nih.gov>  
Contact: [nisc.zoo@nhgri.nih.gov](mailto:nisc.zoo@nhgri.nih.gov)  
----- Project Information  
Center Project name: anu  
Center Clone name: 025E06

The sequence data in this record represents an 'enhanced'

version of a Phase 2 submission. Specifically, the indicated  
order and orientation of each sequence contig has been  
established using one or more of the following: read-pair  
data from individual subclones, overlaps with neighboring  
clones, alignment with available reference sequence (e.g.,  
human), and/or confirmation by PCR testing. In addition,  
the sequence assembly is based on at least 8X average  
coverage in Q20 bases and has been reviewed to rule out  
gross misassemblies, the low-quality ends of sequence  
contigs have been trimmed away, and each base is associated  
with a Phrap-derived quality score.

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 164542 bases at least Q40  
Consensus quality: 164665 bases at least Q30  
Consensus quality: 164684 bases at least Q20  
Insert size: 131000; agarose-fp  
Insert size: 162000; pulse-field-gel  
Insert size: 164685; sum-of-contigs  
Quality coverage: 14.20x in Q20 bases; agarose-fp  
Quality coverage: 11.49x in Q20 bases; pulse-field-gel  
Quality coverage: 11.30x in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submittor.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1. 164685: contig of 164685 bp in length.

# FEATURES

## source

1. .164685  
Location/Qualifiers  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
/clone="RP43-2556"  
/clone\_lib="RP43"  
1. .164685  
/note="assembly\_fragment"

## misc\_feature

clone\_end:SP6  
vector\_side:left  
clone\_end:T7  
vector\_side:right  
1. .86634  
/note="clone overlaps with GenBank Accession Number  
AC103900 clone RP43-161C11 (center project name anv)"

## ORIGIN

Query Match 92.0%; Score 18.4; DB 14; Length 164685;  
Best Local Similarity 95.0%; Pred. No. 2.1e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAGCCAAACAGGAAACCAAC 20

Db 69134 GAACCAACAGGAAACCAAC 69115

## RESULT 13

### AC148263/c

### LOCUS

### DEFINITION

### ACCESSION

### VERSION

### KEYWORDS

### SOURCE

### ORGANISM

### REFERENCE

### AUTHORS

### REFERENCE

### AUTHORS

### REFERENCE

### AUTHORS

### REFERENCE

### AUTHORS

### REFERENCE

### AUTHORS

### REFERENCE

### AUTHORS

### REFERENCE

### AUTHORS

### REFERENCE

### AUTHORS

### REFERENCE

### AUTHORS

### REFERENCE

### AUTHORS

### REFERENCE

### AUTHORS

### REFERENCE

### AUTHORS

### REFERENCE

### AUTHORS

### REFERENCE

### AUTHORS

### REFERENCE

### AUTHORS

### REFERENCE

### AUTHORS

### REFERENCE

### AUTHORS

### REFERENCE

### AUTHORS

### REFERENCE

### AUTHORS

### REFERENCE

### AUTHORS

### REFERENCE

### AUTHORS

### REFERENCE

### AUTHORS

### REFERENCE

### AUTHORS

### REFERENCE

### AUTHORS

The sequence data in this record represents an 'enhanced'  
version of a Phase 2 submission. Specifically, the indicated  
order and orientation of each sequence contig has been  
established using one or more of the following: read-pair  
data from individual subclones, overlaps with neighboring  
clones, alignment with available reference sequence (e.g.,  
human), and/or confirmation by PCR testing. In addition,  
the sequence assembly is based on at least 8X average



coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 185102 bases at least Q40  
Consensus quality: 185329 bases at least Q30  
Consensus quality: 185476 bases at least Q20  
Insert size: 112000; agarose-fp  
Insert size: 185553; sum-of-contigs  
Quality coverage: 17.46x in Q20 bases; agarose-fp  
Quality coverage: 10.54x in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 3 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submittor.  
\* This sequence will be replaced

\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

\* 1 85542: contig of 85542 bp in length  
\* 85543 134357: contig of 48715 bp in length  
\* 134358 134457: gap of unknown length  
\* 134458 185753: contig of 51296 bp in length.

FEATURES

source

1. .185753  
/organism="Callithrix jacchus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9483"  
/clone\_lib="CH259"  
/notes="BAC resource: http://bacpac.chori.org/"

misc\_feature

1. .85542  
/note="assembly\_fragment"  
clone\_end:T7  
vector\_side:left

gap

85543. .85642  
/estimated\_length=unknown  
85643. .134357  
/note="assembly\_fragment"

misc\_feature

88120. .185753  
/notes="clone overlaps with GenBank Accession Number  
AC148192 clone CH259-121123 (center project name fv1)"

gap

134358. .134457  
/estimated\_length=unknown  
134458. .185753  
/notes="assembly\_fragment"  
clone\_end:SP6  
vector\_side:right

ORIGIN

Query Match 92.0%; Score 18.4; DB 14; Length 185753;  
Best Local Similarity 95.0%; Pred. No. 2.1e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAAACCAAC 20

|||||

Db 21798 GAACCAACAGGAAACCAAC 21779

|||||

RESULT 14

AC157650

LOCUS

AC157650 188678 bp DNA linear HTG 24-FEB-2005

DEFINITION

Mus musculus chromosome 14 clone RP24-573N21, WORKING DRAFT

SEQUENCE, 3 unordered pieces.

ACCESSION

AC157650 AC122761

VERSION

AC157650.1 GI:60223364

KEYWORDS

SOURCE

ORGANISM

Mus musculus

(house mouse)

HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridea; Muridae; Murinae; Mus.  
Wilson.R.K.  
The sequence of Mus musculus clone  
2 (bases 1 to 188678)  
Unpublished  
Wilson.R.K.  
Direct Submission  
Submitted (24-FEB-2005) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
On Feb 24, 2005 this sequence version replaced gi:45268893.

----- Genome Center -----

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: http://genome.wustl.edu

Contact: submissions@watson.wustl.edu

----- Project Information -----

Center project name: M\_BB0573N21

Drafting center: WIBR

----- Summary Statistics -----

Sequencing vector: M13; 0%

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 187241 bases at least Q40

Consensus quality: 187611 bases at least Q30

Consensus quality: 187786 bases at least Q20

-----

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 3 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 1119: contig of 1119 bp in length

\* 1120 1219: gap of unknown length

\* 1220 29257: contig of 28038 bp in length

\* 29258 29357: gap of unknown length

\* 29358 188678: contig of 159321 bp in length.

-----

Location/Qualifiers

1. .188678

/organism="Mus musculus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10090"

/chromosome="14"

/clone="RP24-573N21"

1. .1119

/note="assembly\_name:Contig10"

1120. .1219

/estimated\_length=unknown

1220. .29257

/note="assembly\_name:Contig18"

clone\_end:T7

vector\_side:left

29258. .29357

/estimated\_length=unknown

29358. .188678

/note="assembly\_name:Contig19"

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

ORIGIN

Query Match

Best Local Similarity

Matches 19; Conservative

92.0%; Score 18.4; DB 14; Length 188678;

95.0%; Pred. No. 2.1e+02;

Mismatches 1; Indels 0; Gaps 0;

## COMMENT

On Mar 22, 2003 this sequence version replaced gi:28867043.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

QY 1 GAGCCAAACAGGACCAAC 20

Db 185539 GAGCCAAACAGGACCAAC 185558

## RESULT 15

## AC121536

## LOCUS

## DEFINITION

AC121536 192548 bp DNA linear HTG 22-MAR-2003  
 Mus musculus clone RP24-315H19, WORKING DRAFT SEQUENCE, 7 unordered  
 pieces.

## AC121536

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

HTG; HTGS PHASE1; HTGS DRAFT.  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridae; Muridae; Murinae; Mus.

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

1 (bases 1 to 192548)  
 Birren,B., Nusbaum,C. and Lander,E.  
 Mus musculus, clone RP24-315H19  
 Unpublished  
 2 (bases 1 to 192548)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
 Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguelavkiy,L.,  
 Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,  
 Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,  
 Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
 Faro,S., Ferreira,P., Fitzgerald,M., FitzHugh,W., Gage,D.,  
 Galagan,J., Gardyna,S., Ginde,S., Gird,S., Goyette,M., Graham,L.,  
 Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,  
 Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Lacroque,K.,  
 Lamazares,R., Lander,S., Lehoczy,J., Levine,R., Lindblad-Toh,K.,  
 Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N.,  
 Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,  
 Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,  
 Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,  
 O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,  
 Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,  
 Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,  
 Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,  
 Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,  
 Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,  
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
 Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

## Direct Submission

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

Submitted (20-MAY-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 192548)  
 Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,  
 Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,  
 Boguelavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,  
 Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,  
 Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,  
 Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,  
 Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,  
 Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
 Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,  
 Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,  
 Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,  
 Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,  
 Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,  
 O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,  
 Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
 Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,  
 Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,  
 Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,  
 Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,  
 Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

## Direct Submission

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

Submitted (22-MAR-2003) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA

Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L21340  
 Center clone name: 315 H 19  
 ----- Summary Statistics  
 Sequencing vector: plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 189940 bases at least Q40  
 Consensus quality: 190352 bases at least Q30  
 Consensus quality: 190555 bases at least Q20  
 Insert size: 183000; agarose-fp  
 Insert size: 191948; sum-of-contigs  
 Quality coverage: 11.1 in Q20 bases; agarose-fp  
 Quality coverage: 10.6 in Q20 bases; sum-of-contigs  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 7 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 21971: contig of 21971 bp in length  
 \* 21972 22071: gap of 100 bp  
 \* 22072 22794: contig of 723 bp in length  
 \* 22795 22894: gap of 100 bp  
 \* 22895 23551: contig of 657 bp in length  
 \* 23552 23651: gap of 100 bp  
 \* 23652 24677: contig of 1026 bp in length  
 \* 24678 24777: gap of 100 bp  
 \* 24778 47938: contig of 23161 bp in length  
 \* 47939 48038: gap of 100 bp  
 \* 48039 78010: contig of 29972 bp in length  
 \* 78011 78110: gap of 100 bp  
 \* 78111 192548: contig of 114438 bp in length.  
 Location/Qualifiers  
 1..192548  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /clone="RP24-315H19"  
 /clone\_lib="RPCI-24 Male Mouse BAC"  
 misc\_feature 1..21971  
 /note="assembly\_fragment  
 clone end:SP6  
 vector\_side:left"  
 21972..22071  
 /estimated\_length=100  
 22072..22794  
 /note="assembly\_fragment"  
 22795..22894  
 /estimated\_length=100  
 22895..23551  
 /note="assembly\_fragment"  
 23552..23651  
 /estimated\_length=100  
 23652..24677  
 /note="assembly\_fragment"  
 24678..24777  
 /estimated\_length=100  
 24778..47938  
 /note="assembly\_fragment"  
 47939..48038  
 gap

```

misc_feature /estimated_length=100
              48039..78010
              /note="assembly_fragment"
gap           78011..78110
              /estimated_length=100
misc_feature  78111..192548
              /note="assembly_fragment
clone end:T7
vector_side:right"

ORIGIN
Query Match      92.0%; Score 18.4; DB 14; Length 192548;
Best Local Similarity 95.0%; Pred. NO. 2.1e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAGCCAAACAGGACCAAC 20
    |||||
Db 119254 GAGCCAAACAGGACCAAC 119273

RESULT 16
AC154455/c
LOCUS      AC154455      215308 bp      DNA      linear      ROD 11-MAR-2005
DEFINITION Mus musculus BAC clone RP23-185H13 from 14, complete sequence.
ACCESSION  AC154455 AC124995
VERSION     AC154455.2 GI:58082582
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 215308)
AUTHORS     Tomlinson,C. and Bielicki,L.
TITLE       The sequence of Mus musculus BAC clone RP23-185H13
JOURNAL     Unpublished (2001)
REFERENCE   2 (bases 1 to 215308)
AUTHORS     Wilson,R.K.
TITLE       Direct Submission
JOURNAL     Submitted (30-DEC-2004) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3 (bases 1 to 215308)
AUTHORS     Wilson,R.K.
TITLE       Direct Submission
JOURNAL     Submitted (25-JAN-2005) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4 (bases 1 to 215308)
AUTHORS     Wilson,R.K.
TITLE       Direct Submission
JOURNAL     Submitted (11-MAR-2005) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jan 25, 2005 this sequence version replaced gi:56900081.
            ----- Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu
            Contact: submissions@watson.wustl.edu
            ----- Summary Statistics
            -----
            Center project name: M_BA0185H13
            -----

NOTICE:
This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:

```

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

## SOURCE INFORMATION:

The RPCI-23 BAC Library has been constructed by Kazutoyo Osegawa and Minako Tateo in the laboratory of Pieter de Jong (<http://www.chori.org>) from female C57BL/6J mouse kidney and/or brain genomic DNA. The clone and detailed information can be obtained from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

## NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone.

The sequence of AC124995 has been incorporated into AC154455.

FEATURES  
source

Location/Qualifiers  
1..215308  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="14"  
/map="14"  
/clone="RP23-185H13"  
/clone\_lib="RPCI-23"  
84149..84238  
/note="Sequence derived from one plasmid subclone."

unsure

ORIGIN

Query Match 92.0%; Score 18.4; DB 9; Length 215308;  
Best Local Similarity 95.0%; Pred. NO. 2.1e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 1 GAGCCAAACAGGACCAAC 20  
 |||||  
Db 166601 GAGCCAAACAGGACCAAC 166582

RESULT 17

AC121585 195311 bp DNA linear ROD 08-NOV-2003  
LOCUS AC121585 Mus musculus BAC clone RP23-265M4 from 5, complete sequence.  
DEFINITION AC121585  
ACCESSION AC121585  
VERSION AC121585.3 GI:26801335  
KEYWORDS HTG.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 195311)  
AUTHORS Levy,A., Spalding,L., Mangiapanello,L., Haglund,K. and Abbott,A.  
TITLE The sequence of Mus musculus BAC clone RP23-265M4  
JOURNAL Unpublished (2001)  
REFERENCE 2 (bases 1 to 195311)  
AUTHORS Wilson,R.  
TITLE Sequencing of Mus musculus  
JOURNAL Unpublished (2001)  
REFERENCE 3 (bases 1 to 195311)  
AUTHORS McPherson,J.D. and Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (20-MAY-2002) Genome Sequencing Center, 4444 Forest Park  
 Parkway, St. Louis, MO 63108, USA  
REFERENCE 4 (bases 1 to 195311)  
AUTHORS McPherson,J.D. and Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (14-DEC-2002) Genome Sequencing Center, 4444 Forest Park  
 Parkway, St. Louis, MO 63108, USA  
REFERENCE 5 (bases 1 to 195311)  
AUTHORS McPherson,J.D. and Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (14-DEC-2002) Genome Sequencing Center, 4444 Forest Park  
 Parkway, St. Louis, MO 63108, USA

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Parkway, St. Louis, MO 63108, USA  
6 (bases 1 to 195311)  
Wilson.R.  
Direct Submission  
Submitted (08-NOV-2003) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Dec 14, 2002 this sequence version replaced gi:22711832.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu>  
Contact: [submissions@watson.wustl.edu](mailto:submissions@watson.wustl.edu)  
----- Summary Statistics  
Center project name: M\_BA0265M04  
-----

NOTICE: This sequence may not represent the entire insert of this  
clone. It may be shorter because we only sequence overlapping  
clone sections once, or longer because we provide a small overlap  
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by sequence  
from more than one subclone; and the assembly was confirmed by  
restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren,  
Department of Genetics, Washington University, St. Louis MO. For  
additional information about the map position of this sequence, see  
<http://genome.wustl.edu>

SOURCE INFORMATION:

The RPCI-23 BAC library has been constructed by Kazutoyo Osegawa  
and Minako Tateno in the laboratory of Pieter de Jong  
(<http://www.chori.org>) from female C57BL/6J mouse kidney and/or  
brain genomic DNA. The clone and detailed information can be  
obtained from Research Genetics, Inc. (<http://www.resgen.com>) or  
Pieter de Jong and coworkers at <http://www.chori.org>

NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone.

FEATURES  
source

1..195311  
Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosomes="5"  
/map="5"  
/clone="RP23-265M4"  
/clone\_lib="RPCI-23"

unsure  
1..208  
/note="Sequence derived from one plasmid subclone."

repeat\_region  
1..128  
/rpt\_family="L1"  
repeat\_region  
121..541  
/rpt\_family="L1"  
repeat\_region  
558..1110  
/rpt\_family="L1"  
repeat\_region  
1177..2121  
/rpt\_family="L1"  
repeat\_region  
2106..2420  
/rpt\_family="L1"  
repeat\_region  
3447..4729  
/rpt\_family="L1"  
repeat\_region  
6150..6185  
/rpt\_family="L1"  
repeat\_region  
6846..7588  
/rpt\_family="L1"  
repeat\_region  
7611..8013

repeat\_region  
8014..8151  
/rpt\_family="Alu"  
repeat\_region  
8152..8778  
/rpt\_family="L1"  
repeat\_region  
8785..9121  
/rpt\_family="L1"  
repeat\_region  
9122..9658  
/rpt\_family="ERV1"  
repeat\_region  
9674..10335  
/rpt\_family="L1"  
repeat\_region  
10450..10555  
/rpt\_family="Alu"  
repeat\_region  
10833..11215  
/rpt\_family="ERVK"  
repeat\_region  
11216..11777  
/rpt\_family="RMR19B"  
repeat\_region  
11883..12030  
/rpt\_family="Alu"  
repeat\_region  
12107..12234  
/rpt\_family="MaLR"  
repeat\_region  
12381..12451  
/rpt\_family="L2"  
repeat\_region  
12481..12580  
/rpt\_family="L1"  
repeat\_region  
14688..14868  
/rpt\_family="MER1\_type"  
repeat\_region  
21362..23546  
/rpt\_family="L1"  
repeat\_region  
23991..25140  
/rpt\_family="L1"  
repeat\_region  
25650..25777  
/rpt\_family="L1"  
repeat\_region  
26032..26219  
/rpt\_family="L1"  
repeat\_region  
26265..26342  
/rpt\_family="B4"  
repeat\_region  
26343..26514  
/rpt\_family="L1"  
repeat\_region  
27108..27189  
/rpt\_family="L1"  
repeat\_region  
28052..28818  
/rpt\_family="L1"  
repeat\_region  
29913..30061  
/rpt\_family="Alu"  
repeat\_region  
31090..31242  
/rpt\_family="L1"  
repeat\_region  
31492..31699  
/rpt\_family="B2"  
repeat\_region  
33560..33611  
/rpt\_family="ERV1"  
repeat\_region  
34384..34560  
/rpt\_family="L1"  
repeat\_region  
36209..36255  
/rpt\_family="L2"  
repeat\_region  
36316..36507  
/rpt\_family="B2"  
repeat\_region  
36638..36687  
/rpt\_family="ERV1"  
repeat\_region  
36956..37214  
/rpt\_family="B4"  
repeat\_region  
37756..37796  
/rpt\_family="L1"  
repeat\_region  
41971..42041  
/rpt\_family="B4"  
repeat\_region  
42143..42314  
/rpt\_family="B2"  
repeat\_region  
43718..43897  
/rpt\_family="B2"  
complement(43832..43900)  
/product="tRNA-Ser"  
/note="Likely pseudogene (HMM Sc=27.51 / Sec struct  
Sc=-2.17)"

tRNA

```

repeat_region 44574..44667
                /rpt_family="Alu"
repeat_region 4728..47776
                /rpt_family="ERV1"
repeat_region 49266..49848
                /rpt_family="Achoho"
repeat_region 49852..50297
                /rpt_family="B4"
repeat_region 50387..50506
                /rpt_family="Achoho"
repeat_region 50799..50924
                /rpt_family="L1"
repeat_region 50925..51252
                /rpt_family="ERVK"
repeat_region 53754..53937
                /rpt_family="MER1_type"
repeat_region 54032..54171
                /rpt_family="L1"
repeat_region 54206..54388
                /rpt_family="L1"
repeat_region 54573..55014
                /rpt_family="L1"

```

Query Match 90.0%; Score 18; DB 9; Length 195311;

Best Local Similarity 100.0%; Pred. No. 3.3e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAA 18

Db 124435 GAGCCAAACAGGAACCAA 124452

#### RESULT 18

AC087162

LOCUS

DEFINITION AC087162 207588 bp DNA linear HTG 25-MAR-2001  
Mus musculus clone RP23-232H18, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 109  
unordered pieces.

AC087162

VERSION AC087162.3 GI:13446280

KEYWORDS HTG; HTGS PHASE1.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidea; Muridae; Murinae; Mus.

1 (bases 1 to 207588)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Mus musculus, clone RP23-232H18

Unpublished

2 (bases 1 to 207588)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,  
Barna,N., Bastien,V., Bouckhalter,B., Brown,A.,  
Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,  
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,  
Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J.,  
Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hagos,B., Heathford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,  
Jones,C., Karatas,A., Larocque,K., Lamazares,R., Landers,T.,  
Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,  
Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,  
McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mienga,V.,  
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,  
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,  
Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,  
Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,  
Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S., Severy,P.,  
Sougnuez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,  
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,  
Zembek,L., Zimmer,A. and Zody,M.

TITLE

Direct Submission

JOURNAL

Submitted (09-DEC-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 25, 2001 this sequence version replaced GI:13123247.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR  
Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information

Center project name: L11590

Center Clone name: 232\_H\_18  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 109 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 545: contig of 545 bp in length  
\* 546 645: gap of 100 bp  
\* 646 822: contig of 177 bp in length  
\* 823 922: gap of 100 bp  
\* 923 1568: contig of 646 bp in length  
\* 1569 1668: gap of 100 bp  
\* 1669 1971: contig of 303 bp in length  
\* 1972 2071: gap of 100 bp  
\* 2072 3093: contig of 1022 bp in length  
\* 3094 3193: gap of 100 bp  
\* 3194 3721: contig of 528 bp in length  
\* 3722 3821: gap of 100 bp  
\* 3822 4856: contig of 1035 bp in length  
\* 4857 4956: gap of 100 bp  
\* 4957 5946: contig of 990 bp in length  
\* 5947 6046: gap of 100 bp  
\* 6047 6847: contig of 801 bp in length  
\* 6848 6947: gap of 100 bp  
\* 6948 7581: contig of 634 bp in length  
\* 7582 7681: gap of 100 bp  
\* 7682 8400: contig of 719 bp in length  
\* 8401 8500: gap of 100 bp  
\* 8501 9216: contig of 716 bp in length  
\* 9217 9316: gap of 100 bp  
\* 9317 9811: contig of 495 bp in length  
\* 9812 9911: gap of 100 bp  
\* 9912 10586: contig of 675 bp in length  
\* 10587 10686: gap of 100 bp  
\* 10687 11340: contig of 654 bp in length  
\* 11341 11440: gap of 100 bp  
\* 11441 12168: contig of 728 bp in length  
\* 12169 12268: gap of 100 bp  
\* 12269 12935: contig of 667 bp in length  
\* 12936 13035: gap of 100 bp  
\* 13036 13744: contig of 709 bp in length  
\* 13745 13844: gap of 100 bp  
\* 13845 14348: contig of 504 bp in length  
\* 14349 14448: gap of 100 bp  
\* 14449 15080: contig of 632 bp in length  
\* 15081 15180: gap of 100 bp  
\* 15181 15852: contig of 672 bp in length  
\* 15853 15952: gap of 100 bp  
\* 15953 16509: contig of 657 bp in length  
\* 16510 16709: gap of 100 bp  
\* 16710 17846: contig of 1137 bp in length  
\* 17847 17946: gap of 100 bp  
\* 17947 18571: contig of 625 bp in length  
\* 18572 18671: gap of 100 bp  
\* 18672 19518: contig of 847 bp in length  
\* 19519 19618: gap of 100 bp  
\* 19619 20250: contig of 632 bp in length

\* 20251 20350: gap of 100 bp  
\* 20351 21033: contig of 683 bp in length  
\* 21034 21133: gap of 100 bp  
\* 21134 21558: contig of 425 bp in length  
\* 21559 21658: gap of 100 bp  
\* 21659 22346: contig of 688 bp in length  
\* 22347 22446: gap of 100 bp  
\* 22447 23720: contig of 1274 bp in length  
\* 23721 23820: gap of 100 bp  
\* 23821 24539: contig of 719 bp in length  
\* 24540 24639: gap of 100 bp  
\* 24640 25295: contig of 656 bp in length  
\* 25296 25395: gap of 100 bp  
\* 25396 26131: contig of 736 bp in length  
\* 26132 26231: gap of 100 bp  
\* 26232 26944: contig of 713 bp in length  
\* 26945 27044: gap of 100 bp  
\* 27045 27799: contig of 755 bp in length  
\* 27800 27899: gap of 100 bp  
\* 27900 28692: contig of 793 bp in length  
\* 28693 28792: gap of 100 bp  
\* 28793 29888: contig of 1036 bp in length  
\* 29889 31437: contig of 1449 bp in length  
\* 31438 31537: gap of 100 bp  
\* 31538 32595: contig of 1158 bp in length  
\* 32596 32795: gap of 100 bp  
\* 32796 34017: contig of 1222 bp in length  
\* 34018 34117: gap of 100 bp  
\* 34118 35264: contig of 1147 bp in length  
\* 35265 35364: gap of 100 bp  
\* 35365 36141: contig of 777 bp in length  
\* 36142 36241: gap of 100 bp  
\* 36242 37401: contig of 1160 bp in length  
\* 37402 37501: gap of 100 bp  
\* 37502 38944: contig of 1443 bp in length  
\* 38945 39044: gap of 100 bp  
\* 39045 40336: contig of 1132 bp in length  
\* 40337 40336: gap of 100 bp  
\* 40338 41564: contig of 1228 bp in length  
\* 41565 41664: gap of 100 bp  
\* 41665 43110: contig of 1446 bp in length  
\* 43111 43210: gap of 100 bp  
\* 43211 44218: contig of 1008 bp in length  
\* 44219 44318: gap of 100 bp  
\* 44319 45382: contig of 1084 bp in length  
\* 45383 45482: gap of 100 bp  
\* 45483 46150: contig of 668 bp in length  
\* 46151 46250: gap of 100 bp  
\* 46251 47227: contig of 1477 bp in length  
\* 47228 47827: gap of 100 bp  
\* 47828 48840: contig of 1013 bp in length  
\* 48841 48940: gap of 100 bp  
\* 48941 50355: contig of 1415 bp in length  
\* 50356 50455: gap of 100 bp  
\* 50456 51763: contig of 1308 bp in length  
\* 51764 51863: gap of 100 bp  
\* 51864 53343: contig of 1480 bp in length  
\* 53344 53443: gap of 100 bp  
\* 53444 55038: contig of 1595 bp in length  
\* 55039 55138: gap of 100 bp  
\* 55139 56534: contig of 136 bp in length  
\* 56535 56634: gap of 100 bp  
\* 56635 57911: contig of 1277 bp in length  
\* 57912 58011: gap of 100 bp  
\* 58011 58911: contig of 900 bp in length  
\* 58912 59011: gap of 100 bp  
\* 59012 60123: contig of 1112 bp in length  
\* 60124 60223: gap of 100 bp  
\* 60224 61343: contig of 1120 bp in length  
\* 61344 61443: gap of 100 bp  
\* 61444 63346: contig of 1903 bp in length  
\* 63347 63446: gap of 100 bp

\* 63447 64248: contig of 802 bp in length  
\* 64249 64348: gap of 100 bp  
\* 64349 64553: contig of 2105 bp in length  
\* 64554 66553: gap of 100 bp  
\* 66554 67748: contig of 1195 bp in length  
\* 67749 67848: gap of 100 bp  
\* 67849 69558: contig of 1710 bp in length  
\* 69559 69658: gap of 100 bp  
\* 69659 70839: contig of 1181 bp in length  
\* 70840 70939: gap of 100 bp  
\* 70940 71912: contig of 973 bp in length  
\* 71913 72012: gap of 100 bp  
\* 72013 73882: contig of 1870 bp in length  
\* 73883 73982: gap of 100 bp  
\* 73983 75908: contig of 1326 bp in length

Query Match 90.0%; Score 18; DB 14; Length 207588;  
Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AGCCAAACAGGACCAAA 19  
|||||  
DB 63994 AGCCAAACAGGACCAAA 64011

RESULT 19  
AC138101/c  
LOCUS AC138101 211521 bp DNA linear ROD 30-SEP-2003  
DEFINITION Mus musculus chromosome 5, clone RP23-149H20, complete sequence.  
ACCESSION AC138101  
VERSION AC138101.8 GI:34419719  
KEYWORDS HTG.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE  
1 (bases 1 to 211521)  
Birren,B., Nusbaum,C. and Lander,E.  
Mus musculus chromosome 5, clone RP23-149H20  
Unpublished  
2 (bases 1 to 211521)  
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,  
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,  
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,  
Cook,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,  
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,  
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,  
Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J.,  
Mathews,C., McCarthy,M., Meldrum,J., Meneus,L., Mihova,T.,  
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,  
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,  
Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,  
Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,  
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,  
Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,  
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
Direct Submission

TITLE  
JOURNAL  
Submitted (13-DEC-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 211521)  
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,  
Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,  
Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,  
Collymore,A., Cook,A., Cooke,P., Corum,B., DeArelano,K.,  
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,  
Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,  
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,  
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,

Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., MacLean, C.,  
 MacDonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,  
 Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J.,  
 Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,  
 O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,  
 Rachupka, A., Ramaamy, U., Raymond, C., Retta, R., Riese, C., Rogov, P.,  
 Roman, J., Schauer, S., Schnupack, R., Seaman, S., Severy, P., Smith, C.,  
 Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,  
 Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,  
 Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,  
 Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (12-AUG-2003) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 4 bases 1 to 211521  
 Birren, B., Nusbaum, C. and Lander, E.  
 Direct Submission  
 Submitted (03-SEP-2003) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 5 bases 1 to 211521  
 Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,  
 Anderson, M., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,  
 Boguslavsky, L., Boukhalter, B., Camarata, J., Chang, J., Choepel, Y.,  
 Collamore, A., Cook, A., Cooke, P., Corum, B., DeArelano, K., Faro, S.,  
 Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faroo, S.,  
 Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,  
 Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,  
 Hall, J., Horton, L., Hulme, I., Iliev, I., Johnson, R., Jones, C.,  
 Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,  
 Lindblad-Toh, K., Liu, G., Lui, X., Mabbitt, R., MacLean, C.,  
 MacDonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,  
 Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J.,  
 Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,  
 O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,  
 Rachupka, A., Ramaamy, U., Raymond, C., Retta, R., Riese, C., Rogov, P.,  
 Roman, J., Schauer, S., Schnupack, R., Seaman, S., Severy, P., Smith, C.,  
 Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,  
 Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,  
 Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,  
 Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (30-SEP-2003) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Sep 3, 2003 this sequence version replaced gl:33590156.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information  
 Center project name: L22436  
 Center clone name: 149\_H\_20

## FEATURES

## Source

1. 211521  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="5"  
 /map="5"  
 /clone="RP23-149H20"  
 /clone\_lib="RPC1-23 Female Mouse BAC"

## misc\_feature

1. 6  
 /note="clone boundary  
 clone end:SP6  
 site:EcoRI"

## repeat\_region

complement(26..287)  
 /rpt\_family="Charles5"

## repeat\_region

663..794  
 /rpt\_family="MSTC"

804..836  
 /rpt\_family="(TC)n"  
 repeat\_region  
 849..1059  
 /rpt\_family="(TTC)n"  
 repeat\_region  
 1060..1237  
 /rpt\_family="CT-rich"  
 repeat\_region  
 1073..1082  
 /note="single clone coverage"  
 complement(1294..1474)  
 /rpt\_family="B2\_Mm2"  
 repeat\_region  
 complement(1477..1613)  
 /rpt\_family="B1\_MM"  
 repeat\_region  
 1780..1821  
 /rpt\_family="(TA)n"  
 repeat\_region  
 3144..3182  
 /rpt\_family="(TATATG)n"  
 repeat\_region  
 4468..4566  
 /rpt\_family="MIR"  
 repeat\_region  
 complement(6302..6340)  
 /rpt\_family="L3"  
 repeat\_region  
 complement(6364..6535)  
 /rpt\_family="MTE"  
 repeat\_region  
 complement(6564..6679)  
 /rpt\_family="ORR1D"  
 repeat\_region  
 complement(6689..7244)  
 /rpt\_family="ORR1B-int"  
 repeat\_region  
 complement(7265..7409)  
 /rpt\_family="MLT-int"  
 repeat\_region  
 complement(7608..7728)  
 /rpt\_family="MTE"  
 repeat\_region  
 7729..7777  
 /rpt\_family="(TG)n"  
 repeat\_region  
 complement(7784..7986)  
 /rpt\_family="ORR1D"  
 repeat\_region  
 complement(8404..8517)  
 /rpt\_family="MIR"  
 repeat\_region  
 9343..9544  
 /rpt\_family="B3"  
 repeat\_region  
 complement(10005..10182)  
 /rpt\_family="MLT1J2"  
 repeat\_region  
 complement(10259..10378)  
 /rpt\_family="L2"  
 repeat\_region  
 11570..11623  
 /rpt\_family="(TAGA)n"  
 repeat\_region  
 11624..11670  
 /rpt\_family="(TAGA)n"  
 repeat\_region  
 12226..12263  
 /rpt\_family="(CA)n"  
 repeat\_region  
 12755..12894  
 /rpt\_family="B1\_MM"  
 repeat\_region  
 12897..13000  
 /rpt\_family="(TAAA)n"  
 repeat\_region  
 complement(13041..13093)  
 /rpt\_family="RSINE1"  
 repeat\_region  
 13352..13480  
 /rpt\_family="B1F"  
 repeat\_region  
 13481..13501  
 /rpt\_family="(CAAAA)n"  
 repeat\_region  
 13502..13513  
 /rpt\_family="B1F"  
 repeat\_region  
 complement(13639..13738)  
 /rpt\_family="ID\_B1"  
 repeat\_region  
 complement(13739..13856)  
 /rpt\_family="B1\_MM"  
 repeat\_region  
 complement(13857..13919)  
 /rpt\_family="ID\_B1"  
 repeat\_region  
 complement(13985..14117)  
 /rpt\_family="L1ME2"  
 repeat\_region  
 14235..14385  
 /rpt\_family="B3A"  
 repeat\_region  
 complement(14548..14673)  
 /rpt\_family="ID\_B1"  
 repeat\_region  
 15823..15845



[illegible]



repeat\_region complement(4343..4461)  
/rpt\_family="B1\_MM"  
5274..5345  
/rpt\_family="ID5"  
5660..5834  
/rpt\_family="MTE"  
6224..6296  
/rpt\_family="(CGA)n"  
complement(7218..7250)  
/rpt\_family="MIR"  
7759..7812  
/rpt\_family="(TG)n"  
8686..8781  
/rpt\_family="GA-rich"  
8804..8982  
/rpt\_family="G-rich"  
8979..9109  
/rpt\_family="GA-rich"  
9388..9417  
/rpt\_family="(TG)n"  
9764..9900  
/rpt\_family="(CATATA)n"  
10095..10381  
/rpt\_family="B4A"  
10394..10458  
/rpt\_family="(CA)n"  
10459..10492  
/rpt\_family="(GA)n"  
complement(10734..10852)  
/rpt\_family="MIR"  
10872..10904  
/rpt\_family="GA-rich"  
11314..11350  
/rpt\_family="(A)n"  
11572..11727  
/rpt\_family="MLT1A0"  
11729..12020  
/rpt\_family="GA-rich"  
12021..12179  
/rpt\_family="MLT1A0"  
complement(13197..13311)  
/rpt\_family="PBI"  
13390..13413  
/rpt\_family="(TC)n"  
13415..13451  
/rpt\_family="(TG)n"  
complement(13469..13654)  
/rpt\_family="ORR1C"  
13744..13766  
/rpt\_family="(TG)n"  
complement(13803..14042)  
/rpt\_family="B4"  
14089..14111  
/rpt\_family="AT-rich"  
14253..14455  
/rpt\_family="URR1A"  
complement(14859..14975)  
/rpt\_family="MER117"  
complement(15352..15695)  
/rpt\_family="LIM4"  
15866..15886  
/rpt\_family="GC-rich"  
16109..16398  
/rpt\_family="Charlie7"  
18441..18675  
/rpt\_family="Charlie7"  
18231..18322  
/rpt\_family="ID3"  
19202..19423  
/rpt\_family="URR1A"  
20957..21166  
/rpt\_family="ID\_B1"  
complement(21440..21639)

/rpt\_family="LTR67"  
Query Match 90.0%; Score 18; DB 9; Length 240821;  
Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 AGCCAAACAGGAACCAA 19  
DB 78381 AGCCAAACAGGAACCAA 78398  
RESULT 21  
AC004237/c  
LOCUS Homo sapiens chromosome 5, P1 clone 565a12 (LBNL H23), complete  
DEFINITION sequence.  
ACCESSION AC004237 AC000959 L42099 L42100 L42101 L43403 L43404 L43405 L43406  
L49044 L49045 L77041 L77042 L77043  
VERSION AC004237.1 GI:2914666  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
REFERENCE 1 (bases 1 to 38715)  
AUTHORS Connolly,K.S., Gunning,K.M., Davis,C.A., Kadner,K., Subramanian,S.,  
Miguel,T., Lewis,K.D., Fridlyand,J., Alcivare,D., Benke,J.A.,  
Bondoc,M., Bowen,E., Chiang,A., Critz,P., Jaklevic,M.A., Lindo,K.,  
Lindquist,K., Miller,C., Patel,S., Piscia,C., Riley,B.E.,  
Rojeski,H., Sarmiento,R., Yu,C., Montenegro,M., Aerts,A., Chung,A.,  
Abrajaano,A., Baker,M., Gau,C., Jett,J., Ko,C., Beall,K.,  
Woolley,J.P., Stultz,J.L., Kimmerly,W. and Martin,C.H.  
Sequencing of human chromosome 5  
Unpublished  
2 (bases 1 to 38715)  
REFERENCE Ricke,D.O.  
AUTHORS Large Scale Sequence Analysis and Annotation with the Sequence  
Comparison Analysis (SCAN) System  
Unpublished  
3 (bases 1 to 38715)  
AUTHORS Connolly,K.S., Gunning,K.M., Davis,C.A., Kadner,K., Subramanian,S.,  
Miguel,T., Lewis,K.D., Fridlyand,J., Alcivare,D., Benke,J.A.,  
Bondoc,M., Bowen,E., Chiang,A., Critz,P., Jaklevic,M.A., Lindo,K.,  
Lindquist,K., Miller,C., Patel,S., Piscia,C., Riley,B.E.,  
Rojeski,H., Sarmiento,R., Yu,C., Montenegro,M., Aerts,A., Chung,A.,  
Abrajaano,A., Baker,M., Gau,C., Jett,J., Ko,C., Beall,K.,  
Woolley,J.P., Stultz,J.L., Kimmerly,W. and Martin,C.H.  
Direct Submission  
TITLE Submitted (27-FEB-1998) Human Genome Center, DOE Joint Genome  
Institute, Lawrence Berkeley National Laboratory, MS 74-157,  
Berkeley, CA 94720, U.S.A.  
COMMENT Sequence submitted by:  
DOE Joint Genome Institute  
For further information about this sequence, including its location  
and relationship to other sequences, please visit the sequence  
archive Website (<http://www-hgc.lbl.gov/sequence-archive.html>) or  
send email to [human@genome.lbl.gov](mailto:human@genome.lbl.gov)  
This clone maps between the markers D5S642 and D5S1867.  
FEATURES  
Location/Qualifiers  
1..38715  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/map="5q"  
/clone="565a12"  
/note="LBNL H23"  
repeat\_region 30..53  
/note="(AAAG)6"  
/rpt\_type=tandem  
/rpt\_unit="aag"

```
repeat_region complement(1070..1376)
repeat_region /rpt_family="Alu"
1072..1097
/nt="T)26"
/rpt_type=tandem
/rpt_unit="t"
repeat_region complement(1377..1695)
repeat_region /rpt_family="AluSx"
1377..1400
/nt="T)24"
/rpt_type=tandem
/rpt_unit="t"
repeat_region 1808..2128
repeat_region /rpt_family="Alu"
repeat_region complement(2322..2727)
repeat_region /rpt_family="Alu"
2719..2825
/rpt_family="MLT1B"
repeat_region 3119..3238
repeat_region /rpt_family="L1"
3511..3700
/rpt_family="Alu"
repeat_region 3767..4075
repeat_region /rpt_family="L1"
4238..4343
repeat_region /rpt_family="L1"
4289..4598
repeat_region /rpt_family="AluSx"
4530..5006
repeat_region /rpt_family="AluSc"
4809..4830
/nt="(A)22"
/rpt_type=tandem
/rpt_unit="a"
repeat_region 5147..6117
repeat_region /rpt_family="L1"
6317..6512
/rpt_family="MIR"
misc_feature complement(6788..6871)
repeat_region /nt="GRAIL 2 excellent exon, frame 2"
8585..8654
/rpt_family="MIR"
repeat_region 9770..10114
/rpt_family="Alu"
10699..134219
/nt="KIF3"
/nt="Kinesin-like protein KIF3"
join(<10702..10966,18682..18826,18968..19057,24477..24575,
28194..28339,28745..28935,29249..29423,34123..334219)
/nt="KIF3"
/codon_start=1
/evidence=not experimental
/product="KIF3"
/db_xref="GI:2914667"
/translation="INKEPSCDNKVVRCPLNREKSMCYKQAVSDEMRTTI
TVHKDSENPETFTVPFSPKQDVNLTPIDSVLEGTIFAYGQTGKTI
FMVEGVRAIPELRIPISFAHIFGHIAKAGDTRFLVRVSYLEIYNBEVDLLGKQ
TORLEVRPDGVVYIKDLAYVNNADMDRIITLGHKNKRGSGVATNMNHSRSHAI
FTITCESKGIDGNHVRMGLHLVDLAVKEQATGATGQRLKEATKINISLTIGN
VISALVDKSTHVPYRNSKTLRLQSLGNSKTMCMANIGPADYNDETISTLYAN
RAKNIKKARINEDPKALLRFQKEIEELKKLBEGBEISGSDISGSEEDDDDEGEV
GEDGKRRK"
misc_feature 10702..10975
/nt="KIF3"
/nt="GRAIL 2 excellent exon, frame 0"
1162..11270
/rpt_family="L1"
repeat_region complement(11409..11692)
repeat_region /rpt_family="Alu"
11950..12030
/rpt_family="L1"
```

```
repeat_region complement(12242..12686)
repeat_region /rpt_family="Alu"
13173..13470
/rpt_family="Alu"
repeat_region complement(13567..14184)
repeat_region /rpt_family="L1"
complement(14192..15001)
/rpt_family="Alu"
repeat_region complement(15688..15901)
repeat_region /rpt_family="Alu"
complement(15890..16237)
/rpt_family="L1"
repeat_region complement(16234..16378)
repeat_region /rpt_family="Alu"
16407..16434
/nt="(TATT)7"
/rpt_type=tandem
/rpt_unit="tatt"
repeat_region complement(16416..16751)
repeat_region /rpt_family="AluSx"
16736..16985
/rpt_family="Alu"
repeat_region 17877..17977
repeat_region /rpt_family="MIR2"
complement(18211..18287)
/rpt_family="MIR"
repeat_region complement(19546..19622)
repeat_region /rpt_family="MIR"
19682..19753
/rpt_family="MIR"
repeat_region 20324..20373
/nt="(AT)25"
/rpt_type=tandem
/rpt_unit="at"
repeat_region complement(20371..22106)
repeat_region /rpt_family="L1"
25039..25162
/rpt_family="Alu"
repeat_region 25399..25710
/rpt_family="Alu"
repeat_region 26252..26871
/rpt_family="Alu"
repeat_region complement(27700..27811)
misc_feature 28196..28335
/nt="KIF3"
/nt="GRAIL 2 excellent exon, frame 0"
28738..28935
/nt="KIF3"
/nt="GRAIL 2 excellent exon, frame 0"
29288..29423
/nt="KIF3"
/nt="GRAIL 2 excellent exon, frame 1"
29745..29788
/nt="KIF3"
/nt="GRAIL 2 excellent exon, frame 1"
complement(30370..30700)
/rpt_family="L1MB3"
30972..31294
/rpt_family="Alu"
repeat_region 31267..32599
/nt="(A)24"
/rpt_type=tandem
/rpt_unit="a"
repeat_region complement(31300..31599)
repeat_region /rpt_family="L1"
complement(31850..31974)
```

Query Match 87.0%; Score 17.4; DB 8; Length 38715;  
Best Local Similarity 94.7%; Pred. No. 8e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACAAA 19



```

repeat_region      /rpt family="AluY"
                    complement(6297..6393)
repeat_region      /rpt family="L1ME4A"
                    complement(6407..7134)
repeat_region      /rpt family="L1MA2"
                    complement(7152..7388)
repeat_region      /rpt family="MIR"
                    complement(7637..7755)
repeat_region      /rpt family="MIR"
                    complement(7637..7755)
repeat_region      8024..8079
                    /rpt family="MIR"
repeat_region      complement(9718..9813)
                    /rpt family="L1M4"
repeat_region      10892..11008
                    /rpt family="MIR"
repeat_region      complement(11044..11325)
                    /rpt family="AluSx"
repeat_region      complement(11994..12088)
                    /rpt family="L2"
repeat_region      12211..12282
                    /rpt family="L3"
                    12547..12558
                    /note="<30 qual SINGL region"
                    12594..12597
                    /note="<30 qual SINGL region."
                    complement(12803..12968)
                    /rpt family="MER20"
repeat_region      12971..13461
                    /rpt family="MER74A"
repeat_region      complement(13480..13791)
                    /rpt family="AluSc"
repeat_region      complement(13829..13938)
                    /rpt family="MERSA"
                    14356..14467
                    /rpt family="L2"
repeat_region      14486..14522
                    /rpt family="TTAA)n"
repeat_region      15071..15103
                    /rpt family="T-rich"
repeat_region      15933..16027
                    /rpt family="CT-rich"
                    16406..17124
                    /rpt family="LTR1B"
repeat_region      complement(17685..17850)
                    /rpt family="L1MA10"
repeat_region      complement(17862..18161)
                    /rpt family="L1MC1"
                    18616..18868
                    /rpt family="L2"
repeat_region      complement(18877..19070)
                    /rpt family="MER3"
repeat_region      20422..20707
                    /rpt family="HAL1b"
repeat_region      21184..21478
                    /rpt family="AluSp"
repeat_region      21589..22174
                    /rpt family="AluJo"
repeat_region      21784..22088
                    /rpt family="AluY"
repeat_region      22601..22869

Query Match      87.0%; Score 17.4; DB 8; Length 59731;
Best Local Similarity 94.7%; Pred. No. 7.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 AGCCAAACAGGACCAAC 20
Db      32862 AGCCAAACAGGACCAAC 32844

RESULT 23
AP006630/c      AP006630      60126 bp      DNA      linear      PLN 03-FEB-2004
LOCUS

```

```

DEFINITION      Lotus corniculatus var. japonicus genomic DNA, chromosome 2,
                  clone:ljf12f12, TM0028a, complete sequence.
ACCESSION      AP006630.1 GI:41688315
VERSION      HTG.
KEYWORDS
SOURCE
ORGANISM
Lotus corniculatus var. japonicus (Lotus japonicus)
Lotus corniculatus var. japonicus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
Lotus.
REFERENCE
1      Kato,T., Sato,S., Nakamura,Y., Kaneko,T., Asamizu,E. and Tabata,S.
      Structural Analysis of a Lotus japonicus Genome. V. Sequence
      Features and Mapping of Sixty-four TAC Clones Which Cover the 6.4
      Mb Regions of the Genome
      DNA Res. 10, 277-285 (2003)
REFERENCE
2      (bases 1 to 60126)
      Sato,S.
DIRECT SUBMISSION
Submitted (12-NOV-2003) Shusei Sato, Kazusa DNA Research Institute,
Department of Plant Gene Research; 2-6-7 Kazusa-kamatari, Kisarazu,
Chiba 292-0818, Japan (E-mail:ssato@kazusa.or.jp,
URL:http://www.kazusa.or.jp/, Tel:81-438-52-3935(ex.2337),
Fax:81-438-52-3934)
FEATURES
source
1..60126
   /organism="Lotus corniculatus var. japonicus"
   /mol_type="genomic DNA"
   /variety="japonicus"
   /db_xref="taxon:34305"
   /chromosome="2"
   /clone="Lj12f12"
   /clone_lib="LJT library"
   /note="TM0028a, a part of TAC clone:TM0028, synonym:Lotus
japonicus"

ORIGIN
Query Match      87.0%; Score 17.4; DB 15; Length 60126;
Best Local Similarity 94.7%; Pred. No. 7.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 AGCCAAACAGGACCAAC 20
Db      54867 AGCCAAACAGGACCAAC 54849

RESULT 24
AX695578      72732 bp      DNA      linear      PAT 31-MAR-2003
LOCUS
DEFINITION      Sequence 1205 from Patent WO03008583.
ACCESSION      AX695578
VERSION      AX695578.1 GI:29418730
KEYWORDS
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1      Morris,D.W. and Engelhard,B.K.
      Novel compositions and methods for cancer
      Patent: WO 03008583-A 1205 30-JAN-2003;
      Sagres Discovery (US)
FEATURES
source
1..72732
   /organism="Mus musculus"
   /mol_type="unassigned DNA"
   /db_xref="taxon:10090"

ORIGIN
Query Match      87.0%; Score 17.4; DB 6; Length 72732;
Best Local Similarity 94.7%; Pred. No. 7.3e+02;

```

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGCAAAACAGGAACCAAC 20  
 ||||| ||||| |||||  
 Db 6897 AGCAAAACAGGAACCAAC 6915

AC105375 143065 bp DNA linear HTG 08-AUG-2002  
 Felis catus clone RP86-176J15, WORKING DRAFT SEQUENCE, 5 ordered  
 pieces:  
 AC105375  
 HTG; HTGS\_PHASE2; HTGS\_DRAFT.  
 SOURCE Felis catus (cat)  
 ORGANISM Felis catus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Felidae;  
 Felinae; Felis.  
 1 (bases 1 to 143065)  
 Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,  
 Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,  
 Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,  
 Haghghi,P., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E., Laric,P.,  
 Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,  
 Margulies,E.H., Masiello,C., Maskeri,B., Mastrian,S.D.,  
 McCloskey,J.C., McDowell,J., Paguirigan,C., Pearson,R.,  
 Portnoy,M.E., Prasad,A., Schueler,M.G., Stautropop,S., Thomas,J.W.,  
 Thomas,P.J., Touchman,J.W., Tsurgoon,C., Vogt,J.L., Walker,M.A.,  
 Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.  
 NISC Comparative Sequencing Initiative  
 Unpublished

2 (bases 1 to 143065)  
 Green,E.D.  
 Direct Submission  
 Submitted (03-JAN-2002) NIH Intramural Sequencing Center, 8717  
 Grovemont Circle, Gaithersburg, MD 20877, USA  
 3 (bases 1 to 143065)  
 Green,E.D.  
 Direct Submission  
 Submitted (08-AUG-2002) NIH Intramural Sequencing Center, 8717  
 Grovemont Circle, Gaithersburg, MD 20877, USA  
 On Aug 8, 2002 this sequence version replaced gi:18042294.  
 ----- Genome Center  
 Center: NIH Intramural Sequencing Center  
 Center code: NISC  
 Web site: http://www.nisc.nih.gov  
 Contact: nisc.zoo@nhgri.nih.gov  
 ----- Project Information  
 Center project name: awb  
 Center clone name: 176J15

The sequence data in this record represents an 'enhanced'  
 version of a Phase 2 submission. Specifically, the indicated  
 order and orientation of each sequence contig has been  
 established using one or more of the following: read-pair  
 data from individual subclones, overlaps with neighboring  
 clones, alignment with available reference sequence (e.g.,  
 human), and/or confirmation by PCR testing. In addition,  
 the sequence assembly is based on at least 8X average  
 coverage in Q20 bases and has been reviewed to rule out  
 gross misassemblies, the low-quality ends of sequence  
 contigs have been trimmed away, and each base is associated  
 with a Phrap-derived quality score.  
 ----- Summary Statistics  
 Sequencing vector: plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 141388 bases at least Q40  
 Consensus quality: 142081 bases at least Q30  
 Consensus quality: 142489 bases at least Q20  
 Insert size: 128000; agarose-fp

Insert size: 142665; sum-of-contigs  
 Quality coverage: 11.21x in Q20 bases; agarose-fp  
 Quality coverage: 10.06x in Q20 bases; sum-of-contigs  
 -----

\* NOTE: This is a 'working draft' sequence. It currently  
 consists of 5 contigs. Gaps between the contigs  
 are represented as runs of N. The order of the pieces  
 is believed to be correct as given, however the sizes  
 of the gaps between them are based on estimates that have  
 been provided by the submitter.

\* This sequence will be replaced  
 by the finished sequence as soon as it is available and  
 the accession number will be preserved.

\* 1 59848: contig of 59848 bp in length  
 \* 59849 59948: gap of unknown length  
 \* 59949 85977: contig of 25949 bp in length  
 \* 85977 85997: gap of unknown length  
 \* 85998 97089: contig of 11092 bp in length  
 \* 97089 97189: gap of unknown length  
 \* 97189 103765: contig of 6576 bp in length  
 \* 103766 143065: gap of unknown length  
 \* 103766 143065: contig of 39200 bp in length.

## FEATURES

## source

1..143065  
 /organism="Felis catus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9685"  
 /clone="RP86-176J15"  
 /clone\_lib="RP86"  
 1..59848  
 /note="assembly\_fragment  
 clone\_end:T7  
 vector\_side:left"  
 1..57166  
 /note="clone overlaps with GenBank Accession Number  
 AC105374 clone RP86-130P15 (center project name azp)"  
 59849..59948  
 /estimated\_length=unknown  
 59949..85997  
 /note="assembly\_fragment"  
 85998..85997  
 /estimated\_length=unknown  
 85998..97089  
 /note="assembly\_fragment"  
 97090..97189  
 /estimated\_length=unknown  
 97190..103765  
 /note="assembly\_fragment"  
 103766..103865  
 /estimated\_length=unknown  
 103866..143065  
 /note="assembly\_fragment  
 clone\_end:SP6  
 vector\_side:right"  
 135301..143065  
 /note="clone overlaps with GenBank Accession Number  
 AC108194 clone RP86-494M19 (center project name awi)"

## misc\_feature

1..59848  
 /note="assembly\_fragment  
 clone\_end:T7  
 vector\_side:left"  
 1..57166  
 /note="clone overlaps with GenBank Accession Number  
 AC105374 clone RP86-130P15 (center project name azp)"  
 59849..59948  
 /estimated\_length=unknown  
 59949..85997  
 /note="assembly\_fragment"  
 85998..85997  
 /estimated\_length=unknown  
 85998..97089  
 /note="assembly\_fragment"  
 97090..97189  
 /estimated\_length=unknown  
 97190..103765  
 /note="assembly\_fragment"  
 103766..103865  
 /estimated\_length=unknown  
 103866..143065  
 /note="assembly\_fragment  
 clone\_end:SP6  
 vector\_side:right"  
 135301..143065  
 /note="clone overlaps with GenBank Accession Number  
 AC108194 clone RP86-494M19 (center project name awi)"

## misc\_feature

1..57166  
 /note="clone overlaps with GenBank Accession Number  
 AC105374 clone RP86-130P15 (center project name azp)"  
 59849..59948  
 /estimated\_length=unknown  
 59949..85997  
 /note="assembly\_fragment"  
 85998..85997  
 /estimated\_length=unknown  
 85998..97089  
 /note="assembly\_fragment"  
 97090..97189  
 /estimated\_length=unknown  
 97190..103765  
 /note="assembly\_fragment"  
 103766..103865  
 /estimated\_length=unknown  
 103866..143065  
 /note="assembly\_fragment  
 clone\_end:SP6  
 vector\_side:right"  
 135301..143065  
 /note="clone overlaps with GenBank Accession Number  
 AC108194 clone RP86-494M19 (center project name awi)"

## gap

59849..59948  
 /estimated\_length=unknown  
 59949..85997  
 /note="assembly\_fragment"  
 85998..85997  
 /estimated\_length=unknown  
 85998..97089  
 /note="assembly\_fragment"  
 97090..97189  
 /estimated\_length=unknown  
 97190..103765  
 /note="assembly\_fragment"  
 103766..103865  
 /estimated\_length=unknown  
 103866..143065  
 /note="assembly\_fragment  
 clone\_end:SP6  
 vector\_side:right"  
 135301..143065  
 /note="clone overlaps with GenBank Accession Number  
 AC108194 clone RP86-494M19 (center project name awi)"

## misc\_feature

59849..59948  
 /estimated\_length=unknown  
 59949..85997  
 /note="assembly\_fragment"  
 85998..85997  
 /estimated\_length=unknown  
 85998..97089  
 /note="assembly\_fragment"  
 97090..97189  
 /estimated\_length=unknown  
 97190..103765  
 /note="assembly\_fragment"  
 103766..103865  
 /estimated\_length=unknown  
 103866..143065  
 /note="assembly\_fragment  
 clone\_end:SP6  
 vector\_side:right"  
 135301..143065  
 /note="clone overlaps with GenBank Accession Number  
 AC108194 clone RP86-494M19 (center project name awi)"

## misc\_feature

59849..59948  
 /estimated\_length=unknown  
 59949..85997  
 /note="assembly\_fragment"  
 85998..85997  
 /estimated\_length=unknown  
 85998..97089  
 /note="assembly\_fragment"  
 97090..97189  
 /estimated\_length=unknown  
 97190..103765  
 /note="assembly\_fragment"  
 103766..103865  
 /estimated\_length=unknown  
 103866..143065  
 /note="assembly\_fragment  
 clone\_end:SP6  
 vector\_side:right"  
 135301..143065  
 /note="clone overlaps with GenBank Accession Number  
 AC108194 clone RP86-494M19 (center project name awi)"

## misc\_feature

59849..59948  
 /estimated\_length=unknown  
 59949..85997  
 /note="assembly\_fragment"  
 85998..85997  
 /estimated\_length=unknown  
 85998..97089  
 /note="assembly\_fragment"  
 97090..97189  
 /estimated\_length=unknown  
 97190..103765  
 /note="assembly\_fragment"  
 103766..103865  
 /estimated\_length=unknown  
 103866..143065  
 /note="assembly\_fragment  
 clone\_end:SP6  
 vector\_side:right"  
 135301..143065  
 /note="clone overlaps with GenBank Accession Number  
 AC108194 clone RP86-494M19 (center project name awi)"

## misc\_feature

59849..59948  
 /estimated\_length=unknown  
 59949..85997  
 /note="assembly\_fragment"  
 85998..85997  
 /estimated\_length=unknown  
 85998..97089  
 /note="assembly\_fragment"  
 97090..97189  
 /estimated\_length=unknown  
 97190..103765  
 /note="assembly\_fragment"  
 103766..103865  
 /estimated\_length=unknown  
 103866..143065  
 /note="assembly\_fragment  
 clone\_end:SP6  
 vector\_side:right"  
 135301..143065  
 /note="clone overlaps with GenBank Accession Number  
 AC108194 clone RP86-494M19 (center project name awi)"

## misc\_feature

59849..59948  
 /estimated\_length=unknown  
 59949..85997  
 /note="assembly\_fragment"  
 85998..85997  
 /estimated\_length=unknown  
 85998..97089  
 /note="assembly\_fragment"  
 97090..97189  
 /estimated\_length=unknown  
 97190..103765  
 /note="assembly\_fragment"  
 103766..103865  
 /estimated\_length=unknown  
 103866..143065  
 /note="assembly\_fragment  
 clone\_end:SP6  
 vector\_side:right"  
 135301..143065  
 /note="clone overlaps with GenBank Accession Number  
 AC108194 clone RP86-494M19 (center project name awi)"

## misc\_feature

59849..59948  
 /estimated\_length=unknown  
 59949..85997  
 /note="assembly\_fragment"  
 85998..85997  
 /estimated\_length=unknown  
 85998..97089  
 /note="assembly\_fragment"  
 97090..97189  
 /estimated\_length=unknown  
 97190..103765  
 /note="assembly\_fragment"  
 103766..103865  
 /estimated\_length=unknown  
 103866..143065  
 /note="assembly\_fragment  
 clone\_end:SP6  
 vector\_side:right"  
 135301..143065  
 /note="clone overlaps with GenBank Accession Number  
 AC108194 clone RP86-494M19 (center project name awi)"

## ORIGIN

Query Match 87.0%; Score 17.4; DB 14; Length 143065;  
 Best Local Similarity 94.7%; Pred. No. 6.7e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAAA 19  
 ||||| ||||| ||||| |||||

Db 122531 GAGCCAAACAGGAACCAAA 122513

## RESULT 26

AP005010/c AP005010 146936 bp DNA linear HTG 28-MAR-2002  
 DEFINITION Oryza sativa (japonica cultivar-group) chromosome 2 clone P0615D09,  
 \*\*\* SEQUENCING IN PROGRESS \*\*\*.

```

ACCESSION      AP005010
VERSION        AP005010.1  GI:19773550
KEYWORDS       HTG; HTGS PHASE2.
SOURCE         Oryza sativa (japonica cultivar-group)
ORGANISM       Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
               Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
               Ehrhartoideae; Oryzeae; Oryza.
REFERENCE      1
AUTHORS        Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE          Oryza sativa nipponbare (GA3) genomic DNA, chromosome 2, PAC
               clone:P0615D09
JOURNAL        Published Only in Database (2002)
REFERENCE      2 (bases 1 to 146936)
AUTHORS        Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE          Direct Submission
JOURNAL        Submitted (27-MAR-2002) Takuji Sasaki, National Institute of
               Agrobiological Sciences, Rice Genome Research Program; Kannondai
               2-1-2, Tsukuba, Ibaraki 305-8602, Japan
               (E-mail:tsasaki@ias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
               Tel:81-298-38-7411, Fax:81-298-38-7468)
COMMENT        NOTE: It currently consists of 1 contigs. Gaps between the contigs
               are represented as runs of N. The order of the pieces is believed
               to be correct as given, however the sizes of the gaps between them
               are based on estimates that have been provided by the submitter. This
               sequence will be replaced by the finished sequence as soon as it is
               available and the accession number will be preserved.
               * NOTE: This is a 'working draft' sequence.
               * This sequence will be replaced
               * by the finished sequence as soon as it is available and
               * the accession number will be preserved.
FEATURES       source
               Location/Qualifiers
               1..146936
               /organism="Oryza sativa (japonica cultivar-group)"
               /mol_type="genomic DNA"
               /cultivar="Nipponbare"
               /db_xref="taxon:39947"
               /chromosomes="2"
               /clone="P0615D09"
ORIGIN
Query Match      87.0%; Score 17.4; DB 14; Length 146936;
Best Local Similarity 94.7%; Pred. No. 6.7e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GAGCCAAACAGGACCAAA 19
    |||||
Db 44819 GAGCCAAACAGGACCAAA 44801
RESULT 27
AC100737
LOCUS          AC100737 148090 bp DNA linear ROD 11-AUG-2004
DEFINITION    Mus musculus chromosome 7, clone RP24-360I22, complete sequence.
ACCESSION     AC100737
VERSION       AC100737.18 GI:51101173
KEYWORDS      HTG.
SOURCE        Mus musculus (house mouse)
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
               Sciurognathi; Muridae; Murinae; Mus.
REFERENCE     1 (bases 1 to 148090)
AUTHORS       Birren,B., Nusbaum,C. and Lander,E.
TITLE         Mus musculus chromosome 7, clone RP24-360I22
JOURNAL       Unpublished
REFERENCE     2 (bases 1 to 148090)
AUTHORS       Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
               Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
               Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
               Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,A.,
               Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
               Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
               McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlanga,V.,
               Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
               O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
               Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
               Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,
               Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
               Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
               Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
               Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
               Zimmer,A. and Zody,M.
               Direct Submission
               Submitted (11-AUG-2004) Whitehead Institute/MIT Center for Genome
               Research, 320 Charles Street, Cambridge, MA 02141, USA
               All repeats were identified using RepeatMasker:
               Smit, A.F.A. & Green, P. (1996-1997)
               http://ftp.genome.washington.edu/RM/RepeatMasker.html

```

```

TITLE          Direct Submission
JOURNAL        Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
               Research, 320 Charles Street, Cambridge, MA 02141, USA
               3 (bases 1 to 148090)
REFERENCE      Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
               Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
               Bloom,T., Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J.,
               Choepel,Y., Collamore,A., Cooke,A., Cooke,P., Corum,B.,
               DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
               Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D.,
               Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
               Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
               Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,
               Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
               MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,
               McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlanga,V.,
               Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
               O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
               Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
               Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,
               Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
               Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
               Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
               Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
               Zimmer,A. and Zody,M.
               Direct Submission
               Submitted (23-JUN-2004) Whitehead Institute/MIT Center for Genome
               Research, 320 Charles Street, Cambridge, MA 02141, USA
               4 (bases 1 to 148090)
REFERENCE      Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
               Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
               Bloom,T., Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J.,
               Choepel,Y., Collamore,A., Cooke,A., Cooke,P., Corum,B.,
               DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
               Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D.,
               Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
               Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
               Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,
               Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
               MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,
               McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlanga,V.,
               Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
               O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
               Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
               Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,
               Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
               Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
               Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
               Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
               Zimmer,A. and Zody,M.
               Direct Submission
               Submitted (11-AUG-2004) Whitehead Institute/MIT Center for Genome
               Research, 320 Charles Street, Cambridge, MA 02141, USA
               On Aug 11, 2004 this sequence version replaced gi:49065753.
               All repeats were identified using RepeatMasker:
               Smit, A.F.A. & Green, P. (1996-1997)
               http://ftp.genome.washington.edu/RM/RepeatMasker.html

```

----- Genome Center  
 Center: Whitehead Institute/MIT Center for Genome Research  
 Center code: WIBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@broad.mit.edu](mailto:sequence_submissions@broad.mit.edu)  
 ----- Project Information  
 Center project name: L17465  
 Center clone name: 360\_I\_22  
 -----

```

FEATURES
  source      Location/Qualifiers
    1..148090
      /organism="Mus musculus"
      /mol_type="genomic DNA"
      /db_xref="taxon:10090"
      /chromosome="7"
      /map="7"
      /clone="RP24-360I22"
      /clone_lib="RPCI-24 Male Mouse BAC"
      complement(1..1262)
      /rpt_family="L1"
    1..4
      /notes="clone boundary"
      clone end:SP6
      site:MboI
      complement(1254..7491)
      /rpt_family="L1_MM"
      complement(8345..8700)
      /rpt_family="MTD"
      8703..8757
      /rpt_family="MTD"
      8894..9000
      /rpt_family="GA-rich"
      9426..9476
      /rpt_family="(TCC)n"
      10265..10440
      /rpt_family="MLTIL"
      complement(10881..10964)
      /rpt_family="MTD"
      complement(11288..11527)
      /rpt_family="MTB"
      complement(11544..11884)
      /rpt_family="MTE"
      complement(13012..13154)
      /rpt_family="RMER1B"
      complement(13567..13802)
      /rpt_family="RMER1B"
      14047..14078
      /rpt_family="AT-rich"
      15898..15944
      /rpt_family="L2"
      15982..16012
      /rpt_family="AT-rich"
      16169..16209
      /rpt_family="(CAAAA)n"
      complement(16508..16612)
      /rpt_family="ID_B1"
      complement(16636..16753)
      /rpt_family="L1"
      complement(16779..17099)
      /rpt_family="L1"
      18679..18755
      /rpt_family="(TG)n"
      18808..18907
      /rpt_family="(TG)n"
      complement(18910..19169)
      /rpt_family="B4A"
      complement(19170..19260)
      /rpt_family="URR1B"
      19410..19551
      /rpt_family="RSN1"
      19836..20481
      /rpt_family="L1_MM"
      20485..20543
  
```

```

/rpt_family="(GGA)n"
20544..20722
/rpt_family="(GAA)n"
22585..22755
/rpt_family="ID_B1"
22833..23158
/rpt_family="Lx8"
23457..23506
/rpt_family="AT-rich"
complement(23652..23946)
/rpt_family="MLTIC"
complement(24096..24195)
/rpt_family="L2"
24498..25083
/rpt_family="Lx6"
25105..25259
/rpt_family="B3A"
25260..25307
/rpt_family="(TA)n"
25307
/notes="one sub-clone has extra ATAT here"
25308..25319
/rpt_family="B3A"
complement(25311..25349)
/rpt_family="MLT1"
25350..25393
/rpt_family="(CA)n"
complement(25394..25529)
/rpt_family="MLT1"
complement(26998..27120)
/rpt_family="MIR"
complement(28288..29308)
/rpt_family="Lx5"

Query Match      87.0%; Score 17.4; DB 9; Length 148090;
Best Local Similarity 94.7%; Pred.No. 6.7e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy  1 GAGCCAAACAGGACCAAA 19
Db  137777 GAGCCAAACAGGACCAAA 137795

RESULT 28
AC106811/c
LOCUS      AC106811      148362 bp      DNA      linear      PRI 15-FEB-2002
DEFINITION Homo sapiens chromosome 5 clone RP11-551B15, complete sequence.
ACCESSION  AC106811
VERSION     AC106811.2  GI:18677377
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homnidae; Homo.
REFERENCE   1 (bases 1 to 148362)
AUTHORS    DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE      Direct Submission
JOURNAL    Unpublished
REFERENCE   2 (bases 1 to 148362)
AUTHORS    DOE Joint Genome Institute.
TITLE      Direct Submission
JOURNAL    Submitted (12-JAN-2002) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3 (bases 1 to 148362)
AUTHORS    DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE      Direct Submission
JOURNAL    Submitted (15-FEB-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Feb 15, 2002 this sequence version replaced gi:18139361.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
  
```

www-shgc.stanford.edu  
 Quality: Phrap Quality >=40 99.8% of sequence;  
 Estimated Total Number of Errors is 0.6.

## FEATURES

Location/Qualifiers  
 1. .148362  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="5"  
 /clone="RP11-551B15"

## ORIGIN

Query Match 87.0%; Score 17.4; DB 8; Length 148362;  
 Best Local Similarity 94.7%; Pred.No. 6.7e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAA 19  
 ||| ||||| ||||| |||||  
 Db 89749 GAGGCAACAGGACCAA 89731

## RESULT 29

AC117937 152861 bp DNA linear MAM 24-MAR-2004  
 DEFINITION Canis familiaris clone RP81-182F3, complete sequence.  
 ACCESSION AC117937  
 VERSION AC117937.4 GI:45680464  
 KEYWORDS HTG.  
 SOURCE Canis familiaris (dog)  
 ORGANISM Canis familiaris  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;  
 Canis.

## REFERENCE

AUTHORS 1 (bases 1 to 152861)  
 Antonellis,A., Ayele,K., Benjamin,B., Blakesley,R.W.,  
 Bouffard,G.G., Brinkley,C., Brooks,S., Chu,G., Coleman,B.,  
 Coleman,H., Daki,N., Engle,J., Granite,S., Guan,X., Gupta,J.,  
 Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Hu,P., Hurle,B.,  
 Idol,J.R., Jones,C., Karlins,E., Kim,H., Kwong,P., Latic,P.,  
 Larson,S., Lee-Lin,S.-Q., Legaapi,R., Maduro,Q.D., Maduro,V.B.,  
 Margulies,E.H., Masiello,C., Maskeri,B., McDowell,J.,  
 Mullikin,J.C., Paquigan,C., Portnoy,M.E., Prasad,A., Puri,O.,  
 Reddix-Dugue,N., Schandler,K., Schueler,M.G., Shah,K., Sison,C.,  
 Statropop,S., Thomas,J.W., Thomas,P.J., Tsipouri,V., Vogt,J.L.,  
 Wetherby,K.D., Young,A. and Green,E.D.  
 NISC Comparative Sequencing Initiative  
 Unpublished

## TITLE

JOURNAL NISC Comparative Sequencing Initiative  
 REFERENCE 2 (bases 1 to 152861)  
 AUTHORS Green,E.D.  
 TITLE Direct Submission  
 JOURNAL Submitted (11-APR-2002) NIH Intramural Sequencing Center, 8717  
 Government Circle, Gaithersburg, MD 20877, USA

## REFERENCE

AUTHORS 3 (bases 1 to 152861)  
 Green,E.D.  
 TITLE Direct Submission  
 JOURNAL Submitted (01-JUN-2002) NIH Intramural Sequencing Center, 8717  
 Government Circle, Gaithersburg, MD 20877, USA

## REFERENCE

AUTHORS 4 (bases 1 to 152861)  
 Green,E.D.  
 TITLE Direct Submission  
 JOURNAL Submitted (24-MAR-2004) NIH Intramural Sequencing Center, 8717  
 Government Circle, Gaithersburg, MD 20877, USA

## COMMENT

On Mar 24, 2004 this sequence version replaced gi:21306659.  
 ----- Genome Center  
 Center: NIH Intramural Sequencing Center  
 Center code: NISC  
 Web site: <http://www.nisc.nih.gov>  
 Contact: [nisc\\_zoo@hgri.nih.gov](mailto:nisc_zoo@hgri.nih.gov)  
 ----- Project Information  
 Center project name: dgm  
 Center clone name: 182F03

This sequence was finished as follows unless otherwise noted:

all regions were double-stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

CLONE LENGTH: This sequence represents the entire insert of this clone unless otherwise noted. If there are overlapping clones, the overlaps are noted in the beginning and end of the Features section.

## FEATURES

source 1. .152861  
 Location/Qualifiers  
 /organism="Canis familiaris"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9615"  
 /clone="RP81-182F3"  
 /clone\_lib="RP81"  
 /note="BAC resource: <http://bacpac.chori.org/>"  
 misc\_feature 1. .32457  
 /note="low quality single stranded/single chemistry region"  
 misc\_feature 29445. .29495  
 /note="single clone coverage"  
 misc\_feature 29464. .29468  
 /note="low quality single stranded/single chemistry region"  
 misc\_feature 29470. .29472  
 /note="low quality single stranded/single chemistry region"  
 misc\_feature 29474. .29495  
 /note="low quality single stranded/single chemistry region"  
 misc\_feature 29533. .29534  
 /note="low quality single stranded/single chemistry region"  
 misc\_feature 31275. .31280  
 /note="low quality single stranded/single chemistry region"  
 misc\_feature 31335. .31336  
 /note="low quality single stranded/single chemistry region"  
 misc\_feature 32473. .32475  
 /note="low quality single stranded/single chemistry region"  
 misc\_feature 32618  
 /note="low quality single stranded/single chemistry region"  
 misc\_feature 32671. .32672  
 /note="low quality single stranded/single chemistry region"  
 misc\_feature 32675  
 /note="low quality single stranded/single chemistry region"  
 misc\_feature 32696  
 /note="low quality single stranded/single chemistry region"  
 misc\_feature 110389. .110454  
 /note="single clone coverage"  
 misc\_feature 127791. .128156  
 /note="single clone coverage"  
 misc\_feature 141955. .141968  
 /note="single clone coverage"  
 misc\_feature 141965  
 /note="low quality single stranded/single chemistry region"

## ORIGIN

Query Match 87.0%; Score 17.4; DB 4; Length 152861;  
 Best Local Similarity 94.7%; Pred.No. 6.7e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;



```

QY      1 GAGCCAAACGGAACCAAA 19
Db      152229 GAGCCAAACGGAACCAAA 152247

RESULT 30
LOCUS   CR450814/c
DEFINITION Zebrafish DNA sequence from clone CH211-157B11 in linkage group 6, complete sequence.
ACCESSION CR450814
VERSION   CR450814.4 GI:65301311
KEYWORDS  HTG.
SOURCE    Danio rerio (zebrafish)
ORGANISM  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 158142)
AUTHORS   Glithero,R.
TITLE     Direct Submission
JOURNAL   Submitted (14-MAY-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk
COMMENT   On May 14, 2005 this sequence version replaced gi:63147127.
          ----- Genome Center
          Center: Wellcome Trust Sanger Institute
          Center code: SC
          Web site: http://www.sanger.ac.uk
          Contact: zfish-help@sanger.ac.uk
          -----
          During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
          This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.
          The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WormPEP; Information on the WormPEP database can be found at http://www.sanger.ac.uk/projects/C_elegans/wormpep
          Zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.
          Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhirong Bao and Sean Eddy, submitted), and those beginning 'drr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see http://www.sanger.ac.uk/projects/D_rerio/fishmask.shtml
          CH211-157B11 is from a CHORI-211 BAC library
          VECTOR: pTARBAC2.1
          -----
          Location/Qualifiers
            1..158142
              /organism="Danio rerio"
              /mol_type="genomic DNA"
              /db_xref="taxon:7955"
              /clone="CH211-157B11"
              /clone_lib="CHORI-211"

ORIGIN
Query Match      87.0%; Score 17.4; DB 5; Length 158142;
Best Local Similarity 94.7%; Pred. No. 6.6e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 AGCCAAACAGGAACCAAAAC 20
Db      120045 AGCCAAACAGGAATCAAAAC 120027

RESULT 31
LOCUS   AL136218
DEFINITION Human DNA sequence from clone RP11-101D11 on chromosome 13 Contains the 5' end of a novel gene (FLJ12577), the 5' end of the gene for CLIL8 protein (CLLD8) and 2 CpG islands, complete sequence.
ACCESSION AL136218
VERSION   AL136218.26 GI:11120977
KEYWORDS  HTG; CLLD8; CLLL8; CpG island; FLJ12577.
SOURCE    Homo sapiens (human)
ORGANISM  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 159863)
AUTHORS   Pelan,S.
TITLE     Direct Submission
JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
COMMENT   Clone requests: clonerequest@sanger.ac.uk
          On Nov 8, 2000 this sequence version replaced gi:11034476.
          The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WormPEP; Information on the WormPEP database can be found at http://www.sanger.ac.uk/projects/C_elegans/wormpep
          This sequence was generated from part of bacterial clone contigs of human chromosome 13, constructed by the Sanger Centre Chromosome 13 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr13
          -----
          Location/Qualifiers
            1..159863
              /organism="Homo sapiens"
              /mol_type="genomic DNA"
              /db_xref="taxon:9606"
              /chromosome="13"
              /clone="RP11-101D11"
              /clone_lib="RPCI-11.1"
              /note="Clone left end; RP11-101D11"
              join(complement(7252..72642), complement(53919..54060), complement(48086..48250), complement(30833..30951), complement(21863..22031), complement(15298..15357), complement(10796..10861), complement(3066..3209), complement(AL138875.8:69476..71819))
              /gene="RP11-103J18.3"
              /locus_tag="RP11-103J18.3-001"
              join(complement(7252..72642), complement(53919..54060), complement(48086..48250), complement(30833..30951), complement(21863..22031), complement(15298..15357),

FEATURES
source
  1..159863
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /chromosome="13"
    /clone="RP11-101D11"
    /clone_lib="RPCI-11.1"

misc_feature
  1
    /note="Clone left end; RP11-101D11"

gene
  join(complement(7252..72642), complement(53919..54060), complement(48086..48250), complement(30833..30951), complement(21863..22031), complement(15298..15357), complement(10796..10861), complement(3066..3209), complement(AL138875.8:69476..71819))

mRNA
  join(complement(7252..72642), complement(53919..54060), complement(48086..48250), complement(30833..30951), complement(21863..22031), complement(15298..15357),

```

	complement (10796. .10861), complement (3066. .3209), /gene="RP11-103J18.3" /locus_tag="RP11-103J18.3-001" /note="match: ESTs: AA001436 AA234362 AA234412 AA669484 AA716066 AA776711 AI247543 AI252223 AI341575 AI346282 AI417215 AI807383 AI906088 AI956166 AU125107 AW242839 AW345911 AW481483 AW503193 BE042886 BE281207 BE537934 BF224359 BF680686 BF953165 BI337304 BM822014 BM976125 BQ021348 BQ25082 W37890 match: cDNAs: AK022639 BC010993" /locus_tag="RP11-103J18.3-002" /note="complement (53919. .54029), complement (48086. .48250), complement (30833. .30951), complement (21863. .22031), complement (15298. .15357), complement (10796. .10861), complement (3066. .3209), complement (AL138875.8:71640. .71819)" /locus_tag="RP11-103J18.3-001" /standard_name="OTTHUMP00000018410" /note="match: proteins: AAH20570 Q06138 Q21643 Q8VDZ8 Q91WB8 Q91YL0 Q96FG1 Q9DB16 Q9H9S4 Q9Y376" /codon_start=1 /protein_id="CAI10816.1" /db_xref="GI:55859690" /db_xref="InterPro:IPR004892" /translation="MKKPLFSKSHKNPAEIVKILKONLAILEKODKKTDKASEVSK SLQWKEILCGTNEKPPTEAVAQAQLYSSGLLVTLADIQLDFGKKDVTQIFN NLRQIGTRSPTEVYSAPHLFVILKONLAILEKODKKTDKASEVSK FSNQDFPKYVELSTFIADSDAFATKDLTRHKVLVADFLQYNDTIFEDYKLLQ SENYVTKRSLGELILDRHNFATMTKYISKPNLKLMMNLRLDKSPNQFEAFHV KVFVASPHKTQPIVEILLKNQPKLIEFLSSFKERTDDEQFADEKNYLIKQIRLDKK TAP" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus_tag="RP11-103J18.3-002" /locus
--	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

AC009786/c  
LOCUS AC009786 179947 bp DNA linear HTG 23-APR-2000  
DEFINITION Homo sapiens clone RP11-44J9, WORKING DRAFT SEQUENCE, 4 unordered  
pieces.  
ACCESSION AC009786  
VERSION AC009786.2 GI:7637769  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 179947)  
AUTHORS Birren, B., Linton, J., Nussbaum, C., Lander, E., Allen, N., Anderson, M.,  
TITLE Birren, B., Linton, J., Nussbaum, C. and Lander, E.  
JOURNAL Homo sapiens, clone RP11-44J9  
REFERENCE 2 (bases 1 to 179947)  
AUTHORS Birren, B., Linton, J., Nussbaum, C., Lander, E., Allen, N., Anderson, M.,  
Baker, J., Baldwin, J., Barna, N., Beckerly, K., Benn, J., Brown, A.,  
Castle, A., Cerny, J., Colangelo, M., Collins, S., Collymore, A.,  
Cooke, P., DeArelano, K., Depayre, E., Devon, K., Dewar, K.,  
Donelan, L., Doyle, M., Ferreira, P., FitzHugh, W., Forrest, C.,  
Funke, R., Gage, D., Galagan, J., Gardyna, S., Gilbert, D., Grant, G.,  
Hagob, B., Heaford, A., Horton, L., Howland, J. C., Jones, C., Kann, L.,  
Karacas, A., Lehoczy, J., Lieu, C., Locke, K., Macdonald, P.,  
Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J.,  
Meldrim, J., Molla, M., Morris, W., Morrow, J., Mychaleckyl, J.,  
Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Pavlin, B.,  
Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P.,  
Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A.,  
Tesfaye, S., Torruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A.,  
Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.  
Direct Submission  
Submitted (01-SEP-1999) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Apr 23, 2000 this sequence version replaced gi:5815565.  
All repeats were identified using RepeatMasker:  
Smith, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L1076  
Center clone name: 44\_J\_9  
----- Summary Statistics  
Sequencing vector: M13; M77815; 100% of reads  
Chemistry: Dye-primer-amerham; 3% of reads  
Chemistry: Dye-terminator Big Dye; 97% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 177174 bases at least Q40  
Consensus quality: 178432 bases at least Q30  
Consensus quality: 178928 bases at least Q20  
Insert size: 180000; agarose-fp  
Insert size: 179647; sum-of-contigs  
Quality coverage: 7.3 in Q20 bases; agarose-fp  
Quality coverage: 7.3 in Q20 bases.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 1529: contig of 1529 bp in length  
\* 1530: gap of 100 bp  
\* 1630 27837: contig of 26208 bp in length  
\* 27838 27937: gap of 100 bp  
\* 27938 81075: contig of 53138 bp in length  
\* 81076 81175: gap of 100 bp

\* 81176 179947: contig of 98772 bp in length.  
FEATURES  
source  
1. .179947  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="RP11-44J9"  
/clone\_lib="RPC1-11 Human Male BAC"  
misc\_feature  
1. .1529  
/note="assembly\_fragment  
clone\_end:77  
vector\_side:right"  
1530. .1629  
/estimated\_length=100  
1630. .27837  
/note="assembly\_fragment  
clone\_end:SP6  
vector\_side:right"  
27838. .27937  
/estimated\_length=100  
27938. .81075  
/note="assembly\_fragment"  
81075. .81175  
/estimated\_length=100  
81176. .179947  
/note="assembly\_fragment"  
ORIGIN  
Query Match 87.0%; Score 17.4; DB 14; Length 179947;  
Best Local Similarity 94.7%; Pred. No. 6.5e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 1 GAGCCAAACAGGACCAAA 19  
|||||  
DB 58475 GAGCCAAACAGGACCAAA 58457  
  
RESULT 33  
CR925835  
LOCUS CR925835 180180 bp DNA linear HTG 14-MAY-2005  
DEFINITION Danio rerio clone DKEX-86E23, WORKING DRAFT SEQUENCE, 3 unordered  
pieces.  
ACCESSION CR925835  
VERSION CR925835.2 GI:65301274  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
SOURCE Danio rerio (zebrafish)  
ORGANISM Danio rerio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
Cypriniformes; Cyprinidae; Danio.  
McLaren, S.  
Direct Submission  
Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On May 14, 2005 this sequence version replaced gi:56310214.  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: zfish-help@sanger.ac.uk  
----- Project Information  
Center project name: zK86E23  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Chemistry: Dye-terminator; 100% of reads  
Consensus quality: 179347 bases at least Q40  
Consensus quality: 179534 bases at least Q30  
Consensus quality: 179722 bases at least Q20  
Insert size: 179980; sum-of-contigs  
Insert size: 183336; 2.9% error; agarose-fp  
Quality coverage: 10.00x in Q20 bases; sum-of-contigs Quality

coverage: 10.18x in Q20 bases; agarose-fp

-----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 3 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence.  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 125970 126069: gap of 100 bp  
 \* 126070 167545: contig of 41476 bp in length  
 \* 167546 167646: gap of 100 bp  
 \* 167646 180180: contig of 12535 bp in length.

## FEATURES

source  
 1..180180  
 /organism="Danio rerio"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7955"  
 /clone="DKEY-86E23"  
 /clone\_lib="Daniokey"  
 misc\_feature  
 1..125969  
 /note="assembly fragment:00812  
 fragment\_chain:1"  
 misc\_feature  
 126070..167545  
 /note="assembly fragment:00155  
 fragment\_chain:1"  
 misc\_feature  
 167646..180180  
 /note="assembly fragment:00007.0"

## ORIGIN

Query Match 87.0%; Score 17.4; DB 14; Length 180180;  
 Best Local Similarity 94.7%; Pred. No. 6.5e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAAA 19

Db 91053 GAACCAACAGGAACCAAA 91071

## RESULT 34

AC005358/c 184886 bp DNA linear PRI 29-AUG-1998  
 LOCUS Homo sapiens chromosome 17, clone hRPK.746\_E\_8, complete sequence.  
 DEFINITION  
 AC005358  
 AC005358.1 GI:3492889

## KEYWORDS

HTG.

Source Homo sapiens (human)

ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homnidae; Homo.

REFERENCE  
 1 (bases 1 to 184886)

AUTHORS  
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE  
 Homo sapiens chromosome 17, clone hRPK.746\_E\_8

REFERENCE  
 2 (bases 1 to 184886)

AUTHORS  
 Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,  
 Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C.,  
 Brown,A., Castie,A., Cerny,J., Cooke,P., Depayre,E., Devon,K.,  
 Dewar,K., Donelan,L., Ferreira,P., FitzHugh,W., Forrest,C.,  
 Funke,R., Gage,D., Gardyna,S., Geraigery,K., Grant,G., Hagos,B.,  
 Horton,L., Howland,J.C., Jacotot,L., Kann,L., Macdonald,P.,  
 Marquis,N., McEwan,P., Morrow,J., McKernan,K., Meldrim,J.,  
 Molla,M., Morris,W., Morrow,J., Mychaleckyj,J., Nahf,R., Naylor,J.,  
 Niloff,M., O'Connor,T., Pavlin,B., Peterson,K., Riley,R.,  
 Roberts,D., Roy,A., Stange-Thomann,N., Stilwell,J., Stojanovic,N.,  
 Stone,C., Subramanian,A., Torruella-Miller,I., Vassiliev,H., Vo,A.,  
 Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and Zody,M.  
 Direct Submission  
 Submitted (31-JUL-1998) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE  
AUTHORS

3 (bases 1 to 184886)  
 Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,  
 Anderson,M., Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J.,  
 Boutwell,C., Brown,A., Castie,A., Cerny,J., Colangelo,M.,  
 Collins,S., Collymore,A., Cooke,P., Corliss,D., Depayre,E.,  
 Devon,K., Dewar,K., Donelan,L., Ferreira,P., FitzHugh,W.,  
 Forrest,C., Funke,R., Gage,D., Gardyna,S., Geraigery,K., Grant,G.,  
 Hagos,B., Heaford,A., Herena,L., Horton,L., Howland,J.C.,  
 Jacotot,L., Jones,C., Kann,L., Karatas,A., Lehoczy,J.,  
 Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,  
 Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,  
 Nahf,R., Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P.,  
 Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A., Severy,P.,  
 Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,  
 Subramanian,A., Tesfaye,S., Tichovolisky,N., Torruella-Miller,I.,  
 Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,  
 Ye,W.J., Zhao,J. and Zody,M.  
 Direct Submission  
 Submitted (29-AUG-1998) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Aug 29, 1998 this sequence version replaced gi:345231.  
 All repeats were identified using RepeatMasker: Smit, A.F.A. &  
 Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Only the first 184.9 kilobases of this clone are being submitted.  
 The remainder overlaps accession AC005274 (WICGR project L350).

## FEATURES

source  
 Location/Qualifiers  
 1..184886  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="17"  
 /map="17"  
 /clone="hRPK.746\_E\_8"  
 /clone\_lib="RPC1-11 human BAC library"  
 complement(229..1257)  
 /rpt\_family="LI"  
 repeat\_region  
 complement(1258..1282)  
 /rpt\_family="(CA)n"  
 repeat\_region  
 complement(1283..1899)  
 /rpt\_family="LI"  
 repeat\_region  
 complement(1900..1919)  
 /rpt\_family="(CAAA)n"  
 repeat\_region  
 complement(1920..2157)  
 /rpt\_family="LI"  
 repeat\_region  
 complement(2158..2458)  
 /rpt\_family="AluYa5"  
 repeat\_region  
 complement(2459..2791)  
 /rpt\_family="LI"  
 repeat\_region  
 complement(2792..2814)  
 /rpt\_family="(TAAA)n"  
 repeat\_region  
 complement(2815..3245)  
 /rpt\_family="LI"  
 repeat\_region  
 3249..3299  
 /rpt\_family="Tigger1"  
 repeat\_region  
 complement(3418..4216)  
 /rpt\_family="LIMD"  
 repeat\_region  
 complement(4220..4513)  
 /rpt\_family="AluX"  
 repeat\_region  
 complement(4552..4856)  
 /rpt\_family="LIMD"  
 repeat\_region  
 complement(4857..5154)  
 /rpt\_family="AluSp"  
 repeat\_region  
 complement(5155..5761)  
 /rpt\_family="LIMD"  
 repeat\_region  
 5914..5977  
 /rpt\_family="AT\_rich"  
 repeat\_region  
 6346..6606  
 /rpt\_family="LMC/D"  
 repeat\_region  
 6837..6902  
 /rpt\_family="AT\_rich"  
 repeat\_region  
 complement(6934..7008)

```
repeat_region /rpt_family="MLR1J"
repeat_region /rpt_family="MIR"
repeat_region 8403..8540
repeat_region /rpt_family="MIR"
repeat_region 9099..9517
repeat_region /rpt_family="L1M4"
repeat_region 9565..9587
repeat_region /rpt_family="AT_rich"
repeat_region 9760..10067
repeat_region /rpt_family="AluSc"
repeat_region 10472..10773
repeat_region /rpt_family="L1MD3"
repeat_region 10804..11106
repeat_region /rpt_family="AluSx"
repeat_region complement(12541..12756)
repeat_region /rpt_family="L2"
repeat_region 12757..12917
repeat_region /rpt_family="MLR1B"
repeat_region complement(12918..13217)
repeat_region /rpt_family="AluSx"
repeat_region 13218..13452
repeat_region /rpt_family="MLR1B"
repeat_region complement(13453..13576)
repeat_region /rpt_family="L2"
repeat_region 13612..13671
repeat_region /rpt_family="MER91C"
repeat_region complement(13672..13852)
repeat_region /rpt_family="Tigger1"
repeat_region 13853..14149
repeat_region /rpt_family="AluSx"
repeat_region complement(14150..15415)
repeat_region /rpt_family="Tigger1"
repeat_region complement(15416..15456)
repeat_region /rpt_family="(TAAA)n"
repeat_region complement(15458..15744)
repeat_region /rpt_family="AluSg"
repeat_region complement(15745..16500)
repeat_region /rpt_family="Tigger1"
repeat_region 16501..16799
repeat_region /rpt_family="AluSg"
repeat_region complement(16800..17032)
repeat_region /rpt_family="Tigger1"
repeat_region 17033..17078
repeat_region /rpt_family="MER91C"
repeat_region complement(17131..17413)
repeat_region /rpt_family="L2"
repeat_region 17536..17819
repeat_region /rpt_family="AluSx"
repeat_region 17820..17850
repeat_region /rpt_family="(CAAAA)n"
repeat_region 17851..17864
repeat_region /rpt_family="AluSx"
repeat_region complement(18059..18111)
repeat_region /rpt_family="L2"
repeat_region complement(20221..20327)
repeat_region /rpt_family="MER5A"
repeat_region 21060..21248
repeat_region /rpt_family="MIR"
repeat_region 21249..21361
repeat_region /rpt_family="MER81"
repeat_region 21362..21405
repeat_region /rpt_family="MIR"
repeat_region complement(21469..21753)
repeat_region /rpt_family="AluSp"
repeat_region complement(22556..22682)
repeat_region /rpt_family="MER94"
repeat_region 24774..25006
repeat_region /rpt_family="L2"
repeat_region complement(25715..25734)
repeat_region /rpt_family="(TAAA)n"
repeat_region complement(25735..26883)
repeat_region /rpt_family="L1PA7"

complement(26884..27228)
/rpt_family="MER82"
27229..27431
/rpt_family="L1MCS"
complement(28245..28279)
/rpt_family="(CA)n"
complement(28950..29539)
/rpt_family="L1MC1"
29549..29628
/rpt_family="L2"
30603..30647
/rpt_family="AT_rich"
complement(30696..30930)
/rpt_family="Charlies"
complement(30934..31088)
/rpt_family="Charlies"
complement(31126..31275)
/rpt_family="L2"
32366..32835
/rpt_family="L1MA8"
33033..33202
/rpt_family="AluSg/x"
33203..33297
/rpt_family="(GGGGA)n"

Query Match 87.0%; Score 17.4; DB 8; Length 184886;
Best Local Similarity 94.7%; Pred. No. 6.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGCCAAACAGGAACCAAC 20
||||| ||||| ||||| |||||
Db 6268 AGCCAAAGGAACCAAC 6250

RESULT 35
AC161229
LOCUS
DEFINITION HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
PROGRESS ***
ACCESSION AC161229
VERSION AC161229.9 GI:72534952
KEYWORDS HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
Birren,B., Nusbaum,C. and Lander,E.
1 (bases 1 to 189104)
Mus musculus chromosome 7, clone RP24-88J19
Unpublished
REFERENCE
2 (bases 1 to 189104)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Baatien,V.,
Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J.,
Choepe,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,
DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
Erickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D.,
Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,
Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
Toplam,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
Zimmer,A. and Zody,M.
```

**TITLE**  
JOURNAL

**REFERENCE**  
AUTHORS

Direct Submission  
Submitted (10-MAY-2005) Broad Institute of MIT and Harvard, 320  
Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 189104)  
Birren,B., Nusbaum,C., Lander,B., Abouelleil,A., Allen,N.,  
Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,  
Bloom,T., Boguslavskiy,L., Boukigalter,B., Camarata,J., Chang,J.,  
Chapel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,...  
DeRellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,  
Erickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D.,  
Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,  
Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,  
Johnson,R., Jones,C., Katat,A., Karatas,A., Kells,C., Landers,T.,  
Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,  
Maclean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,  
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mienga,V.,  
Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,  
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,  
Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,  
Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupack,R.,  
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Stubbs,M., Talamas,J., Testaye,S., Theodore,J.,  
Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,  
Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,  
Zimmer,A. and Zody,M.

**TITLE**  
JOURNAL

Direct Submission  
Submitted (13-AUG-2005) Broad Institute of MIT and Harvard, 320  
Charles Street, Cambridge, MA 02141, USA  
On Aug 13, 2005 this sequence version replaced gi:72255659.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
Center: Broad Institute of MIT and Harvard  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@broad.mit.edu](mailto:sequence_submissions@broad.mit.edu)  
----- Project Information  
Center project name: L32353  
Center clone name: 88\_J\_19  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* been provided by the submittor.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 189104: contig of 189104 bp in length.

**FEATURES**  
source  
1. .189104  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="7"  
/map="7"  
/clone="RP24-88J19"  
/clone\_lib="RP24-88J19"

**ORIGIN**

Query Match 87.0%; Score 17.4; DB 14; Length 189104;  
Best Local Similarity 94.7%; Pred. No. 6.5e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 GAGCCAAACAGGACCAAA 19  
|||||  
Db 35840 GAGCCAAACAGGACCAAA 35858

RESULT 36  
CR847954/c

**LOCUS**  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

CR847954  
Danio rerio chromosome 6 clone DXEY-265K7, WORKING DRAFT SEQUENCE,  
4 unordered pieces.  
CR847954.3 GI:68688253  
HTG; HTGS PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
Danio rerio (zebrafish)  
Danio rerio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
Cypriniformes; Cyprinidae; Danio.  
1 (bases 1 to 189984)  
McLaren,S.  
Direct Submission  
Submitted (02-JUL-2005) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
[zfish-help@sanger.ac.uk](mailto:zfish-help@sanger.ac.uk) Clone requests:  
[http://www.sanger.ac.uk/Projects/D\\_rerio/fags.shtml#dataeight](http://www.sanger.ac.uk/Projects/D_rerio/fags.shtml#dataeight)  
On Jul 5, 2005 this sequence version replaced gi:54260947.

----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: [zfish-help@sanger.ac.uk](mailto:zfish-help@sanger.ac.uk)  
----- Project Information  
Center project name: zK265K7  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Chemistry: Dye-terminator; 100% of reads  
Consensus quality: 188294 bases at least Q40  
Consensus quality: 188403 bases at least Q30  
Consensus quality: 188625 bases at least Q20  
Insert size: 189684; sum-of-contigs  
Quality size: 194086; 1.1% error; agarose-fp  
Quality coverage: 8.54x in Q20 bases; sum-of-contigs Quality  
coverage: 8.41x in Q20 bases; agarose-fp  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 25261: contig of 25261 bp in length  
\* 25262 25361: gap of 100 bp  
\* 25362 51724: contig of 26363 bp in length  
\* 51725 51824: gap of 100 bp  
\* 51825 107899: contig of 56075 bp in length  
\* 107900 107999: gap of 100 bp  
\* 189984: contig of 81985 bp in length.

**FEATURES**  
Location/Qualifiers  
1. .189984  
/organism="Danio rerio"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7955"  
/chromosome="6"  
/clone="DXEY-265K7"  
/clone\_lib="Daniokey"  
1. .25261  
/note="assembly fragment:00013  
fragment\_chain:1"  
25362. .51724  
/note="assembly fragment:00361  
fragment\_chain:1"  
51825. .107899  
/note="assembly fragment:00729  
fragment\_chain:1"  
108000. .189984  
/note="assembly fragment:01515.0"

misc\_feature  
misc\_feature  
misc\_feature  
misc\_feature

ORIGIN

```

Query Match      87.0%; Score 17.4; DB 14; Length 189984;
Best Local Similarity 94.7%; Pred. No. 6.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 AGCCAAACAGGAACCAAAAC 20
      ||||| ||||| ||||| |||||
DB      5493 AGCCAAACAGGATCAAAAC 5475

RESULT 37
AC141646      196787 bp      DNA      linear      HTG 02-SEP-2004
LOCUS      Mus musculus chromosome 9 clone RP23-84E4, WORKING DRAFT SEQUENCE,
DEFINITION      8 unordered pieces.
ACCESSION      AC141646.3 GI:51854801
VERSION      HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ACTIVEFIN.
KEYWORDS      Mus musculus (house mouse)
SOURCE      Mus musculus
ORGANISM      Mus musculus
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
TITLE      Sciurognathi; Muroidae; Muridae; Murinae; Mus.
JOURNAL      1 (bases 1 to 196787)
AUTHORS      Wilson,R.K.
TITLE      The sequence of Mus musculus clone
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 196787)
AUTHORS      McPherson,J.D. and Waterston,R.H.
TITLE      Direct Submission
JOURNAL      Submitted (18-MAR-2003) Genome Sequencing Center, 4444 Forest Park
REFERENCE      Parkway, St. Louis, MO 63108, USA
AUTHORS      3 (bases 1 to 196787)
TITLE      Wilson,R.K.
JOURNAL      Direct Submission
REFERENCE      Submitted (02-SEP-2004) Genome Sequencing Center, 4444 Forest Park
AUTHORS      Parkway, St. Louis, MO 63108, USA
TITLE      On Sep 2, 2004 this sequence version replaced gi:29124326.
JOURNAL      COMMENT

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@wustl.edu
----- Project Information -----
Center project name: M BA0084E04
----- Summary Statistics -----
Sequencing vector: M13; 0%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 193183 bases at least Q40
Consensus quality: 193776 bases at least Q30
Consensus quality: 193991 bases at least Q20
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1372: contig of 1372 bp in length
* 1373 1472: gap of unknown length
* 1473 3599: contig of 2127 bp in length
* 3600 3699: gap of unknown length
* 3700 18151: contig of 14452 bp in length
* 18152 18251: gap of unknown length
* 18252 76024: contig of 57773 bp in length
* 76025 76124: gap of unknown length
* 76125 132788: contig of 56664 bp in length

```

```

FEATURES
source      1..196787
            /organism="Mus musculus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10090"
            /chromosome="9"
            /clone="RP23-84E4"
            /name="Contig21"
            /notes="assembly_name:Contig21"
            /estimated_length=unknown
            /notes="assembly_name:Contig22"
            /estimated_length=unknown
            /notes="assembly_name:Contig23"
            /estimated_length=unknown
            /notes="assembly_name:Contig24"
            /estimated_length=unknown
            /notes="assembly_name:Contig25"
            /estimated_length=unknown
            /notes="assembly_name:Contig26"
            /estimated_length=unknown
            /notes="assembly_name:Contig4"
            /estimated_length=unknown
            /notes="assembly_name:Contig5"

ORIGIN
Query Match      87.0%; Score 17.4; DB 14; Length 196787;
Best Local Similarity 94.7%; Pred. No. 6.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 AGCCAAACAGGAACCAAAAC 20
      ||||| ||||| ||||| |||||
DB      38449 AGCCAAACAGGAACCAAAAC 38467

RESULT 38
CR376854      200904 bp      DNA      linear      VRT 21-JUN-2005
LOCUS      Zebrafish DNA sequence from clone DKEY-197C15 in linkage group 24,
DEFINITION      complete sequence.
ACCESSION      CR376854
VERSION      CR376854.11 GI:69051167
KEYWORDS      HTG.
SOURCE      Danio rerio (zebrafish)
ORGANISM      Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
            1 (bases 1 to 200904)
            Clark,S.
            Direct Submission
            Submitted (18-JUN-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests:

```



COMMENT [http://www.sanger.ac.uk/Projects/D\\_rerio/faq/faq.shtml#dataeight](http://www.sanger.ac.uk/Projects/D_rerio/faq/faq.shtml#dataeight)  
 On Jun 20, 2005 this sequence version replaced gi:67513890.  
 ----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: <http://www.sanger.ac.uk>  
 Contact: [zfish-help@sanger.ac.uk](mailto:zfish-help@sanger.ac.uk)  
 -----  
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.  
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:  
 Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep)  
 Zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.  
 Any regions longer than 1kb tagged as misc-feature 'unsure' are part of a tandem repeat of more than 10kb in length where it has not been possible to anchor the base differences between repeat copies. The region has been built up based on the repeat element to match the total size of repeat indicated by restriction digest, but repeat copies may not be in the correct order and the usual finishing criteria may not apply. DKEY-197C15 is from a Zebrafish BAC library  
 VECTOR: pIndigoBAC-5.  
 Location/Qualifiers  
 1..200904  
 /organism="Danio rerio"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7955"  
 /chromosome="24"  
 /clone="DKEY-197C15"  
 /clone\_lib="DanioKey"  
 ORIGIN  
 Query Match 87.0%; Score 17.4; DB 5; Length 200904;  
 Best Local Similarity 94.7%; Pred. No. 6.4e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GAGCCAAACAGGAACCAAA 19  
 Db 170818 GAACCAACAGGAACCAAA 170836  
 RESULT 39  
 AC155199/c  
 LOCUS AC155199 204615 bp DNA linear HTG 10-FEB-2005  
 DEFINITION Callithrix jacchus clone CH259-246F3, WORKING DRAFT SEQUENCE, 17 unordered pieces.  
 AC155199  
 AC155199.2 GI:58866108  
 VERSION HTG; HTGS PHASE1; HTGS DRAFT.  
 KEYWORDS  
 SOURCE Callithrix jacchus (white-tufted-ear marmoset)  
 ORGANISM Callithrix jacchus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Platyrrhini; Callitrichidae; Callithrix.  
 1 (bases 1 to 204615)  
 REFERENCE  
 AUTHORS Antonellis A., Ayele, K., Benjamin, B., Blakesley, R. W., Boakye, A., Bouffard, G. G., Brinkley, C., Brooks, S., Chu, G., Coleman, H., Engle, J., Gestole, M., Guan, X., Gupta, J., Gutierrez, P., Haghghi, P., Han, J., Hansen, N., Ho, S.-L., Hu, P., Hunter, G., Hurle, B., Idol, J. R., Jones, C., Kwong, P., Laric, P., Larson, S., Lee-Lin, S.-Q., Legaspi, R., Madden, M., Maduro, Q. L., Maduro, V. B., Margulies, E. H., Masiello, C., Maskeri, B., McDowell, J., Mullikin, J. C., Oestreich, J. S., Park, M.,

Portnov, M. E., Prasad, A., Puri, O., Reddix-Dugue, N., Rosas, B., Schandler, K., Schueler, M. G., Sison, C., Stantropop, S., Stephen, E., Taye, A., Thomas, J. W., Thomas, P. J., Tsipouri, V., Vogt, J. L., Wetherby, K. D., Young, A. and Green, E. D.  
 NISC Comparative Sequencing Initiative  
 Unpublished  
 2 (bases 1 to 204615)  
 Green, E. D.  
 Direct Submission  
 TITLE Direct Submission  
 JOURNAL Submitted (12-JAN-2005) NIH Intramural Sequencing Center, 5625 Fishers Lane, Rockville, MD 20852, USA  
 3 (bases 1 to 204615)  
 Green, E. D.  
 Direct Submission  
 TITLE Direct Submission  
 JOURNAL Submitted (10-FEB-2005) NIH Intramural Sequencing Center, 5625 Fishers Lane, Rockville, MD 20852, USA  
 On Feb 10, 2005 this sequence version replaced gi:57544649.  
 ----- Genome Center  
 Center: NIH Intramural Sequencing Center  
 Center code: NISC  
 Web site: <http://www.nisc.nih.gov>  
 Contact: [nisc\\_zoo@nhgri.nih.gov](mailto:nisc_zoo@nhgri.nih.gov)  
 ----- Project Information  
 Center project name: bnf  
 Center clone name: 246P03  
 ----- Summary Statistics  
 Sequencing vector: plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 194068 bases at least Q40  
 Consensus quality: 195836 bases at least Q30  
 Consensus quality: 197075 bases at least Q20  
 Insert size: 175000; agarose-fp  
 Quality coverage: 10.26x in Q20 bases; agarose-fp  
 Quality coverage: 8.84x in Q20 bases; sum-of-contigs  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently consists of 17 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.  
 \* 1 2477: contig of 2477 bp in length  
 \* 2478 2577: gap of unknown length  
 \* 2578 5619: contig of 3042 bp in length  
 \* 5620 5719: gap of unknown length  
 \* 5720 8256: contig of 2537 bp in length  
 \* 8257 8356: gap of unknown length  
 \* 8357 10451: contig of 2095 bp in length  
 \* 10452 10551: gap of unknown length  
 \* 10552 12802: contig of 2251 bp in length  
 \* 12803 12902: gap of unknown length  
 \* 12903 15348: contig of 2446 bp in length  
 \* 15349 15448: gap of unknown length  
 \* 15449 19502: contig of 4054 bp in length  
 \* 19503 19602: gap of unknown length  
 \* 19603 25084: contig of 5482 bp in length  
 \* 25085 25184: gap of unknown length  
 \* 25185 30518: contig of 5334 bp in length  
 \* 30519 30618: gap of unknown length  
 \* 30619 35336: contig of 4818 bp in length  
 \* 35337 35536: gap of unknown length  
 \* 35537 49774: contig of 14238 bp in length  
 \* 49775 49874: gap of unknown length  
 \* 49875 59442: contig of 9568 bp in length  
 \* 59443 59542: gap of unknown length  
 \* 59543 72168: contig of 12626 bp in length  
 \* 72169 72268: gap of unknown length  
 \* 72269 98049: contig of 25781 bp in length  
 \* 98050 98149: gap of unknown length



\* 98150 128011: contig of 29862 bp in length  
 \* 128012 128111: gap of unknown length  
 \* 128112 161695: contig of 33584 bp in length  
 \* 161696 161795: gap of unknown length  
 \* 161796 204615: contig of 42820 bp in length.

## FEATURES

```

source      Location/Qualifiers
            1..204615
             /organism="Callithrix jacchus"
             /mol_type="genomic DNA"
             /db_xref="taxon:9483"
             /clone="CH259-246P3"
             /clone_lib="CH259"
             /notes="BAC resource: http://bacpac.chori.org/"
misc_feature 1..2477
            /notes="assembly_fragment"
            2478..2577
             /estimated_length=unknown
            2578..5619
             /notes="assembly_fragment"
            5620..5719
             /estimated_length=unknown
            5720..8256
             /notes="assembly_fragment"
            8257..8356
             /estimated_length=unknown
            8357..10451
             /notes="assembly_fragment"
            10452..10551
             /estimated_length=unknown
            10552..12802
             /notes="assembly_fragment"
            12803..12902
             /estimated_length=unknown
            12903..15348
             /notes="assembly_fragment"
            15349..15448
             /estimated_length=unknown
            15449..19502
             /notes="assembly_fragment"
            19503..19602
             /estimated_length=unknown
            19603..25084
             /notes="assembly_fragment"
            25085..25184
             /estimated_length=unknown
            25185..30518
             /notes="assembly_fragment"
            30519..30618
             /estimated_length=unknown
            30619..35436
             /notes="assembly_fragment"
            35437..35536
             /estimated_length=unknown
            35537..49774
             /notes="assembly_fragment"
            49775..49874
             /estimated_length=unknown
            49875..59442
             /notes="assembly_fragment"
            59443..59542
             /estimated_length=unknown
            59543..72168
             /notes="assembly_fragment"
            72169..72268
             /estimated_length=unknown
            72269..98049
             /notes="assembly_fragment"
            98050..98149
             /estimated_length=unknown
            98150..128011
             /notes="assembly_fragment"
            128012..128111
             /estimated_length=unknown

```

```

misc_feature 128112..161695
            /notes="assembly_fragment"
            clone end:SP6
            vector side:right"
            161696..161795
             /estimated_length=unknown
            161796..204615
             /note="assembly_fragment"
             clone end:T7
             vector_side:left"

ORIGIN

Query Match      87.0%; Score 17.4; DB 14; Length 204615;
Best Local Similarity 94.7%; Pred. No. 6.4e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2 AGCCACAGGACCAAC 20
Db      1185 AGCCACAGGACCAAC 1167

RESULT 40
AC111278/c
LOCUS      AC111278
DEFINITION Rattus norvegicus clone CH230-50P4, WORKING DRAFT SEQUENCE, 3
ACCESION   AC111278
VERSION    AC111278.4 GI:30578688
KEYWORDS   HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE     Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
            Chordata; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridea; Muridae; Murinae; Rattus.

REFERENCE   1 (bases 1 to 219316)
AUTHORS    Muzny, D., Marie, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J.,
            Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
            Anyalebechi, V., Ayodeji, A., Ayodeji, M., Baca, E., Baden, H.,
            Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
            Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
            Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
            Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
            Chacko, J., Chavez, D., Chen, R., Chen, Y., Chen, Z., Chu, J.,
            Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
            Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
            Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
            Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
            Egan, A., Escotto, M., Eugene, C., Evans, C., Falls, T., Fan, G.,
            Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
            Fraser, C., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
            Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
            Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
            Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
            Hernandez, R., Hines, S., Hladun, S., Hodgson, A., Hogues, M.,
            Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
            Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
            Karpathy, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
            Kowis, C., Kraft, C., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
            Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
            Lorensuhewa, L., Loulaseg, H., Lozado, R., Lu, X., Ma, J.,
            Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
            Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
            Mawhiney, S., McLeod, M., McNeill, T., Meenen, E.,
            Milosavljevic, A., Miner, G., Minja, E., Montanayor, J., Moore, S.,
            Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Naif, L.,
            Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
            Nwaokemele, O., Okwuonu, G., Olarnpungoon, A., Pal, S., Parks, K.,
            Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
            Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.,
            Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M., Reigh, R.,
            Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
            Rives, C., Rodkey, T., Rojas, M., Rose, R., Rose, R., Ruiz, S.,
            Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,

```

Shetty J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D.,  
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,  
 Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,  
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K.,  
 Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,  
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
 Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K.,  
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
 Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,  
 Weinstock, G. and Gibbs, R.A.

# TITLE

JOURNAL

## REFERENCE

AUTHORS

TITLE

JOURNAL

## REFERENCE

AUTHORS

TITLE

JOURNAL

## COMMENT

gap /db\_xref="taxon:10116"  
 gap /clone="CH230-50P4"  
 gap 214138. .214237  
 gap /estimated\_length=unknown  
 gap 216687. .216786  
 gap /estimated\_length=unknown

## ORIGIN

Query Match 87.0%; Score 17.4; DB 14; Length 219316;  
 Best Local Similarity 94.7%; Pred. No. 6.4e-02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGCCAAACAGGAACCAAC 20

Db 166904 AGCCAAAGGAACCAAC 166886

Search completed: November 20, 2005, 18:33:32  
 Job time : 739.099 secs

Direct Submission  
 2 (bases 1 to 219316)  
 Worley, K.C.  
 Direct Submission  
 Submitted (19-FEB-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 219316)  
 Direct Submission  
 Submitted (13-MAY-2003) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On May 13, 2003 this sequence version replaced gi:22857391.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.

## ----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GLXB

Center clone name: CH230-50P4

----- Summary Statistics

Assembly program: Atlas 3.0;

Consensus quality: 209758 bases at least Q40

Consensus quality: 212663 bases at least Q30

Consensus quality: 213965 bases at least Q20

Estimated insert size: 224924; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length  
 \* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 3 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence.  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 214137: contig of 214137 bp in length

\* 214138 214237: gap of unknown length

\* 214238 216686: contig of 2449 bp in length

\* 216687 216786: gap of unknown length

\* 216787 219316: contig of 2530 bp in length.

Location/Qualifiers

1. .219316

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

## FEATURES

SOURCE

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2005, 15:03:55 ; Search time 324.938 Seconds  
(without alignments)  
410.213 Million cell updates/sec

Title: US-10-627-757-22

Perfect score: 20

Sequence: 1 gagccaaacaggaaacaaac 20

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_21.\*

- 1: Geneseqn1980s.\*
- 2: Geneseqn1990s.\*
- 3: Geneseqn2000s.\*
- 4: Geneseqn2001as.\*
- 5: Geneseqn2001bs.\*
- 6: Geneseqn2002as.\*
- 7: Geneseqn2002bs.\*
- 8: Geneseqn2003as.\*
- 9: Geneseqn2003bs.\*
- 10: Geneseqn2003cs.\*
- 11: Geneseqn2003ds.\*
- 12: Geneseqn2004as.\*
- 13: Geneseqn2004bs.\*
- 14: Geneseqn2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	20	ADL14970	ADL14970 Human gla
2	20	100.0	20	ADW86592	ADW86592 PCR prime
3	20	100.0	1074	ADL14953	ADL14953 Human gla
4	20	100.0	1074	ADW86575	ADW86575 Human opt
5	20	100.0	46951	ADL13891	ADL13891 Human opt
6	17.4	87.0	2224	ABZ36233	ABZ36233 Human sec
7	17.4	87.0	72732	9 ADA02687	ADA02687 Mouse Fli
8	17.4	87.0	72732	10 ADB72425	ADB72425 Mouse Fli
9	17.4	87.0	72732	10 ADB72425	ADB72425 Mouse Fli
10	17.4	87.0	110000	10 ADG70447_2	Continuation (3 of
11	17.4	87.0	110000	10 ABZ79565_2	Continuation (3 of
12	17	85.0	263852	13 ADG99460	ADG99460 Murine ki
13	16.8	84.0	483	10 ABZ40103	ABZ40103 N. gonorr
14	16.8	84.0	2285	12 ADH18912	ADH18912 Human cel
15	16.8	84.0	2509	12 ADH18919	ADH18919 Human cel
16	16.8	84.0	2742	13 ADL19326	ADL19326 Human int
17	16.8	84.0	2780	11 ADL131813	ADL131813 Human CDN
18	16.8	84.0	2780	13 ADS83880	ADS83880 Human lym
19	16.8	84.0	2798	6 ABL65835	ABL65835 Lung canc

C 20	16.8	84.0	2798	7 ADS99921	ADS99921 Human int
C 21	16.8	84.0	2798	10 ABX70638	ABX70638 Human CDN
C 22	16.8	84.0	2798	12 ADP10372	ADP10372 Reference
C 23	16.8	84.0	2798	14 ADY19537	ADY19537 DNA encod
C 24	16.8	84.0	2798	14 ADY15613	ADY15613 DNA encod
C 25	16.8	84.0	2800	13 ACN43200	ACN43200 Human dia
C 26	16.8	84.0	2835	6 AAS94871	AAS94871 Human DNA
C 27	16.8	84.0	3244	13 ACN43199	ACN43199 Human dia
C 28	16.8	84.0	31766	6 AAD22781	AAD22781 Human sul
C 29	16.8	84.0	31766	6 AAL50687	AAL50687 Human sul
C 30	16.4	82.0	1071	12 ADJ35017	ADJ35017 DNA encod
C 31	16.4	82.0	1794	8 ADA70575	ADA70575 Rice gene
C 32	16.4	82.0	1794	11 ACL28714	ACL28714 Rice abio
C 33	16.4	82.0	2476	14 ADW10224	ADW10224 Colon pro
C 34	16.4	82.0	95484	12 ADQ97298	ADQ97298 Mouse can
C 35	16.4	82.0	310268	13 ABD32548	ABD32548 Human can
C 36	16	80.0	871	3 AAC50252	AAC50252 Arabidops
C 37	16	80.0	5735	4 ABL23613	ABL23613 Drosophil
C 38	16	80.0	8954	4 ABL23612	ABL23612 Drosophil
C 39	16	80.0	110000	14 AEB39175_01	Continuation (2 of
C 40	16	80.0	184368	14 AEB35722	AEB35722 L. pneumo
C 41	15.8	79.0	23	14 ADZ11811	ADZ11811 Canine ca
C 42	15.8	79.0	177	13 ACF88064	ACF88064 Human SIR
C 43	15.8	79.0	219	2 AAV09870	AAV09870 Human cat
C 44	15.8	79.0	228	14 AEB53070	AEB53070 Human pro
C 45	15.8	79.0	242	12 ACH90106	ACH90106 Human gen

ALIGNMENTS

RESULT 1

ADL14970

ID ADL14970 standard; DNA; 20 BP.

AC ADL14970;

DT 06-MAY-2004 (first entry)

DE Human glaucoma-related optineurin (OPTN) exon 7 PCR primer SR7.

XX Human; glaucoma; optineurin; OPTN; diagnosis; PCR; primer; ss.

OS Homo sapiens.

XX EP1388590-A2.

XX 11-FEB-2004.

XX 29-JUL-2003; 2003EP-00447201.

XX 02-AUG-2002; 2002JP-00226612.

XX (SYSM-) SYSMEX CORP.

XX Kouchi Y, Masago A, Takahata T;

XX WPI; 2004-146134/15.

PT Gene assay for predicting future onset of glaucoma, particularly primary open angle glaucoma or normal ocular tension glaucoma, comprises detecting a mutation of at least one base of the optineurin gene.

XX Claim 9; SEQ ID NO 22; 31pp; English.

XX The present sequence is that of PCR primer SR7 for exon 7 ADL14953 of the glaucoma-associated gene, OPTN (Optineurin) ADL14949. The invention relates to a gene assay method for predicting future onset of primary open angle glaucoma and/or normal ocular tension glaucoma. This involves detecting a mutation in the OPTN gene coding sequence, specifically a substitution of G for A at position 619 and/or a substitution of A for G at position 898 of the OPTN coding sequence. The mutation(s) is detected using a nucleic acid amplification method using primers specific for the

CC different exons of the coding sequence, including primers SR7 and SR7  
 CC ADL14969 for exon 7.

XX SQ Sequence 20 BP; 10 A; 6 C; 4 G; 0 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 20; DB 12; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAAC 20  
 |||||  
 Db 1 GAGCCAAACAGGAACCAAC 20

RESULT 2  
 ADW86592  
 ID ADW86592 standard; DNA; 20 BP.

XX AC ADW86592;

XX 21-APR-2005 (first entry)

XX PCR primer used to amplify human optineurin (OPTN) gene exon 7 Seq 22.

XX glaucoma; optineurin; ophthalmological; ss; ocular disease;  
 KW DNA amplification; genetic marker; PCR; primer.

XX OS Homo sapiens.

XX JP2005034112-A.

XX 10-FEB-2005.

XX 29-JUL-2003; 2003JP-00281897.

XX 02-AUG-2002; 2002JP-00226612.

XX 30-JUN-2003; 2003JP-00188070.

XX (TOAI-) TOA IYO DENSHI KK.

XX WPI; 2005-156038/17.

XX Estimating risk of onset of glaucoma, involves analyzing mutation in any  
 PT one portion of Optic new phosphorous gene, and utilizing analyzed  
 PT mutation as index for estimating risk of onset of glaucoma.

XX Example 1; SEQ ID NO 22; 13pp; Japanese.

XX This invention relates to a novel method for estimating the risk of onset  
 CC of glaucoma. Specifically, it refers to a method that involves analyzing  
 CC a mutation in any one portion of an optineurin (OPTN) gene and utilizing  
 CC the analyzed mutation as an index for estimating the risk of onset of  
 CC glaucoma. The present invention describes oligonucleotides to detect  
 CC mutations that hybridize with one or more portions of the OPTN glaucoma  
 CC related gene. Accordingly, it provides oligos that detect A619G and G898A  
 CC mutations in the human OPTN gene of patients at risk of developing  
 CC glaucoma, in particular primary open-angle glaucoma and/or normal tension  
 CC glaucoma. These mutations will not be present in a sample obtained from a  
 CC healthy person who is not at risk of onset of glaucoma. This  
 CC oligonucleotide sequence is a PCR primer used to amplify a human OPTN  
 CC exon of the invention.

XX SQ Sequence 20 BP; 10 A; 6 C; 4 G; 0 T; 0 U; 0 Other;

Query Match 100.0%; Score 20; DB 14; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAAC 20  
 |||||  
 Db 1 GAGCCAAACAGGAACCAAC 20

RESULT 3  
 ADL14953/C  
 ID ADL14953 standard; DNA; 1074 BP.  
 XX AC ADL14953;  
 XX 06-MAY-2004 (first entry)  
 DT Human glaucoma-related optineurin (OPTN) exon 7.  
 DE Human; glaucoma; optineurin; OPTN; diagnosis; ds.  
 KW Homo sapiens.

XX FH Key Location/Qualifiers  
 FT primer\_bind complement(261..280)  
 FT /\*tag= a  
 FT /\*note= "Primer SP7"  
 FT exon 501..574  
 FT /\*tag= b  
 FT /number= 7  
 FT primer\_bind 765..784  
 FT /\*tag= c  
 FT /\*note= "Primer SR7"

XX EP1388590-A2.

XX 11-FEB-2004.

XX 29-JUL-2003; 2003EP-00447201.

XX 02-AUG-2002; 2002JP-00226612.

XX (SYSM-) SYSMEX CORP.

XX Kouchi Y, Masago A, Takahata T;

XX WPI; 2004-146134/15.

XX Gene assay for predicting future onset of glaucoma, particularly primary  
 PT open angle glaucoma or normal ocular tension glaucoma, comprises  
 PT detecting a mutation of at least one base of the optineurin gene.

XX Example 1; SEQ ID NO 5; 31pp; English.

XX The present sequence comprises exon 7 of the glaucoma-associated gene,  
 CC OPTN (optineurin) ADL14949. The invention relates to a gene assay method  
 CC for predicting future onset of primary open angle glaucoma and/or normal  
 CC ocular tension glaucoma. This involves detecting a mutation in the OPTN  
 CC gene coding sequence, specifically a substitution of G for A at position  
 CC 619 and/or a substitution of A for G at position 898 of the OPTN coding  
 CC sequence. The mutation(s) is detected using a nucleic acid amplification  
 CC method using primers specific for the different exons of the coding  
 CC sequence, including primers SF7 ADL14969 and SR7 ADL14970 for exon 7.

XX SQ Sequence 1074 BP; 279 A; 222 C; 249 G; 324 T; 0 U; 0 Other;

Query Match 100.0%; Score 20; DB 12; Length 1074;  
 Best Local Similarity 100.0%; Pred. No. 18;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAAC 20  
 |||||  
 Db 784 GAGCCAAACAGGAACCAAC 765

RESULT 4

ADW86575/C

ID ADW86575 standard; DNA; 1074 BP.

XX AC ADW86575;

XX 21-APR-2005 (first entry)

```
XX Human optineurin (OPRN) gene exon 7 Seq 5.
DE
XX
XX glaucoma; optineurin; ophthalmological; ds; ocular disease;
KW DNA amplification; genetic marker.
XX
XX Homo sapiens.
OS
XX JP2005034112-A.
PN
XX 10-FEB-2005.
PD
XX
XX 29-JUL-2003; 2003JP-00281897.
PF
XX 02-AUG-2002; 2002JP-00226612.
PR
XX 30-JUN-2003; 2003JP-00188070.
PR
XX (TOAI-) TOA IYO DENSHI KK.
PA
XX WPI; 2005-156038/17.
XX
DR
XX
XX Estimating risk of onset of glaucoma, involves analyzing mutation in any
PT one portion of Optic new phosphorous gene, and utilizing analyzed
PT mutation as index for estimating risk of onset of glaucoma.
XX
XX
PS Example 1; SEQ ID NO 5; 13pp; Japanese.
XX
XX This invention relates to a novel method for estimating the risk of onset
CC of glaucoma. Specifically, it refers to a method that involves analyzing
CC a mutation in any one portion of an optineurin (OPRN) gene and utilizing
CC the analyzed mutation as an index for estimating the risk of onset of
CC glaucoma. The present invention describes oligonucleotides to detect
CC mutations that hybridize with one or more portions of the OPRN glaucoma
CC related gene. Accordingly, it provides oligos that detect A619G and G98A
CC mutations in the human OPRN gene of patients at risk of developing
CC glaucoma, in particular primary open-angle glaucoma and/or normal tension
CC glaucoma. These mutations will not be present in a sample obtained from a
CC healthy person who is not at risk of onset of glaucoma. This
CC polynucleotide sequence is a human OPRN exon of the invention.
XX
XX Sequence 1074 BP; 279 A; 222 C; 249 G; 324 T; 0 U; 0 Other;
SQ
Query Match 100.0%; Score 20; DB 14; Length 1074;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 20; 'Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAGCCAAACAGGACCAAC 20
Db 784 GAGCCAAACAGGACCAAC 765
RESULT 5
ADE13891/c
ID ADE13891 standard; DNA; 46951 BP.
XX
XX ADE13891;
AC
XX 29-JAN-2004 (first entry)
DT
XX Human optineurin gene.
DE
XX Human; optineurin; ds; gene; ophthalmological;
KW single nucleotide polymorphism; SNP; glaucoma;
KW progressive ocular hypertensive disorder; glaucoma related disorder.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH variation replace(391,G)
FT /*tag= a
FT /*standard_name= "Single nucleotide polymorphism"
FT replace(691,G)
FT /*tag= b
```

```
FT variation
FT /standard_name= "Single nucleotide polymorphism"
FT replace(709,G)
FT /*tag= c
FT /standard_name= "Single nucleotide polymorphism"
FT replace(887,A)
FT /*tag= d
FT /standard_name= "Single nucleotide polymorphism"
FT replace(894,T)
FT /*tag= e
FT /standard_name= "Single nucleotide polymorphism"
FT replace(987,C)
FT /*tag= f
FT /standard_name= "Single nucleotide polymorphism"
FT replace(1112,C)
FT /*tag= g
FT /standard_name= "Single nucleotide polymorphism"
FT replace(1505,CC)
FT /*tag= h
FT /standard_name= "Single nucleotide polymorphism"
FT replace(1606,A)
FT /*tag= i
FT /standard_name= "Single nucleotide polymorphism"
FT replace(2405,T)
FT /*tag= j
FT /standard_name= "Single nucleotide polymorphism"
FT replace(2606,G)
FT /*tag= k
FT /standard_name= "Single nucleotide polymorphism"
FT replace(3313,A)
FT /*tag= l
FT /standard_name= "Single nucleotide polymorphism"
FT replace(3555,TT)
FT /*tag= m
FT /standard_name= "Single nucleotide polymorphism"
FT replace(3825,G)
FT /*tag= n
FT /standard_name= "Single nucleotide polymorphism"
FT replace(3629,C)
FT /*tag= o
FT /standard_name= "Single nucleotide polymorphism"
FT replace(3882,TT)
FT /*tag= p
FT /standard_name= "Single nucleotide polymorphism"
FT replace(3988,T)
FT /*tag= q
FT /standard_name= "Single nucleotide polymorphism"
FT replace(4452,A)
FT /*tag= r
FT /standard_name= "Single nucleotide polymorphism"
XX
XX US2003190617-A1.
PN
XX
XX 09-OCT-2003.
PD
XX
XX 06-MAR-2002; 2002US-00091281.
PF
XX
XX 06-MAR-2002; 2002US-00091281.
PR
XX
XX (SIEE/) SI E.
PA (RAYN/) RAYMOND V.
PA (MORI/) MORISSETTE J.
XX
XX Raymond V, Morissette J, Si E;
PI
XX WPI; 2003-864168/80.
XX
XX New nucleic acid sequences of the optineurin gene are useful to detect
PT polymorphisms particularly single nucleotide polymorphisms in the
PT optineurin promoter to diagnose, prognose and treat glaucoma and related
PT disorders.
XX
XX Disclosure; SEQ ID NO 2; 159pp; English.
XX
```

CC The invention relates to an isolated nucleic acid (N1) comprising at  
 CC least 20 but not more than 1500 consecutive nucleotides of the optineurin  
 CC promoter appearing as ADE13890. Also included are the optineurin promoter  
 CC operably linked to a heterologous nucleic acid, a nucleic acid capable of  
 CC detecting a single nucleotide polymorphism (SNP) in the optineurin  
 CC heterologous sequence, diagnosing or prognosing glaucoma in a sample  
 CC obtained from a cell or bodily fluid (comprising glaucoma in a sample  
 CC in a promoter region of the optineurin gene, associated with a glaucoma  
 CC phenotype), detecting a SNP sequence variation in a sample containing  
 CC DNA, detecting the presence of an optineurin promoter sequence variation  
 CC in a sample containing DNA, determining the presence or increased  
 CC susceptibility to glaucoma or to a progressive ocular hypertensive  
 CC disorder resulting in loss of visual field in a patient for the severity  
 CC or progression of glaucoma in a patient, comprising providing  
 CC amplification reaction primers that direct amplification of a selected  
 CC nucleic acid region containing the variation within the optineurin  
 CC promoter and amplifying the DNA) and detecting a polymorphism (comprising  
 CC obtaining a sample containing human genomic DNA, providing a nucleic acid  
 CC capable of detecting a SNP located within an optineurin promoter, and  
 CC detecting the polymorphism). The invention is used to diagnose and  
 CC prognose glaucoma and also to treat glaucoma related disorders. The  
 CC present sequence is the optineurin gene.

XX SQ Sequence 46951 BP; 12703 A; 10108 C; 10051 G; 14089 T; 0 U; 0 Other;

Query Match 100.0%; Score 20; DB 10; Length 46951;

Best Local Similarity 100.0%; Pred. No. 26;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGACCAAAAC 20

|||||

DB 21394 GAGCCAAACAGGACCAAAAC 21375

RESULT 6

ID ABZ36233/c

XX ID ABZ36233 standard; cDNA; 2224 BP.

XX AC ABZ36233;

XX

DT 10-FEB-2003 (first entry)

XX Human secretory polynucleotide SPTM SEQ ID NO 397.

DE Human secretory polynucleotide SPTM SEQ ID NO 397.

XX Human; SPTM; autoimmune disorder; inflammatory disorder; AIDS; anaemia;

KW asthma; Crohn's disease; neurological disorder; epilepsy; cancer;

KW Huntington's disease; Alzheimer's disease; Creutzfeldt-Jakob disease;

KW multiple sclerosis; Parkinson's disease; cell proliferative disorder;

KW anti-inflammatory; immunosuppressive; neuroprotective; neurotropic;

KW neuroleptic; anticonvulsant; cytostatic; antiparkinsonian; anxiolytic;

KW antipsoriatic; antianaemic; anti-HIV; human immunodeficiency virus;

KW secretory polynucleotide; secretory protein; gene; ss.

XX Homo sapiens.

OS

XX WO200283876-A2.

PN

XX 24-OCT-2002.

PD

XX 27-MAR-2002; 2002WO-US009921.

PF

XX 29-MAR-2001; 2001US-0280067P.

PR 29-MAR-2001; 2001US-0280068P.

PR 16-MAY-2001; 2001US-0291280P.

PR 17-MAY-2001; 2001US-0291829P.

PR 17-MAY-2001; 2001US-0291849P.

PR 19-JUN-2001; 2001US-0293428P.

PR 20-JUN-2001; 2001US-0299776P.

PR 20-JUN-2001; 2001US-0300001P.

XX (INCY-) INCYTE GENOMICS INC.

PA

XX

PI Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;  
 PI DuFour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Ameshey SR;  
 PI Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Ameshey SR;  
 PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;  
 PI Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;  
 XX WPI; 2003-075543/07.  
 DR P-PSDB; ABP75791.  
 XX New human secretory proteins and polynucleotides, useful for diagnosing,  
 PT treating or preventing autoimmune/inflammatory disorders (e.g. AIDS),  
 PT neurological disorders (e.g. Alzheimer's), or cell proliferations or  
 PT cancers.

XX Claim 1; SEQ ID NO 397; 458pp + Sequence Listing; English.

XX The invention relates to a secretory polynucleotide (designated sptm)  
 CC comprising any of 567 polynucleotide sequences (ABZ35837-ABZ36403), a  
 CC naturally occurring polynucleotide sequence at least 90 % identical to  
 CC the polynucleotide sequence, a polynucleotide complementary to them or an  
 CC RNA equivalent of them. The polypeptide or polynucleotide are useful for  
 CC treating, preventing or diagnosing a disease or condition associated with  
 CC the expression of functional SPTM. These are particularly useful for  
 CC diagnosing, treating or preventing autoimmune/inflammatory disorders  
 CC (e.g. acquired immunodeficiency syndrome, anaemia, asthma or Crohn's  
 CC disease), neurological disorders (e.g. epilepsy, Huntington's disease,  
 CC dementia, stroke, Alzheimer's disease, Creutzfeldt-Jakob disease,  
 CC multiple sclerosis, cerebral palsy, Parkinson's disease, anxiety,  
 CC schizophrenia or amnesia), or cell proliferative disorders (e.g.  
 CC psoriasis, polycythemia vera, or cancers including adenocarcinoma,  
 CC leukaemia, lymphoma, melanoma, myeloma, sarcoma or cancers of the brain,  
 CC breast, cervix or prostate). Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 2224 BP; 537 A; 551 C; 581 G; 555 T; 0 U; 0 Other;

Query Match 87.0%; Score 17.4; DB 8; Length 2224;

Best Local Similarity 94.7%; Pred. No. 3.2e+02;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGCCAAACAGGACCAAAAC 20

|||||

DB 2042 AACCAACAGGACCAAAAC 2024

RESULT 7

ADA02687

ID ADA02687 standard; DNA; 72732 BP.

XX

AC ADA02687;

XX

DT 06-NOV-2003 (first entry)

XX

DE Mouse Fli1 carcinoma associated gene, SEQ ID NO:1205.

XX

KW Mouse; murine; carcinoma associated; oncogene; carcinoma; cancer; breast;

KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;

KW gene; ds.

XX Mus sp.

XX WO20003057146-A2.

PN

XX 17-JUL-2003.

PD

XX 26-DEC-2002; 2002WO-US041414.

PR

XX 26-DEC-2001; 2001US-00035832.

PR

XX (SAGR-) SAGRES DISCOVERY.

PA

XX Morris DW;

PI

XX WPI; 2003-587068/55.  
 XX New recombinant nucleic acid encoding carcinoma associated protein,  
 PT useful for preparing compositions for treating carcinomas.  
 XX  
 PS Claim 1; SEQ ID NO 1205; 245pp; English.  
 XX  
 CC The invention relates to recombinant carcinoma associated (CA) nucleic  
 CC acid sequences from mouse and human (ADA01482-ADA03094), and to  
 CC recombinant carcinoma associated proteins (CAP) encoded by them. The  
 CC invention also encompasses expression vectors and host cells comprising a  
 CC CA nucleic acid, a polypeptide (especially an antibody) that specifically  
 CC binds to the protein, and a biochip comprising CA nucleic acid or  
 CC fragments thereof. The sequences of the invention were identified using  
 CC oncogenic retroviruses, which insert into the genome of the host organism  
 CC at random. Many of these do not carry transduced host oncogenes or  
 CC pathogenic trans-acting viral genes, meaning that cancer incidence is a  
 CC direct consequence of the effects of proviral integration into host  
 CC protooncogenes. The CA nucleic acid sequences can be used to diagnose  
 CC carcinoma (especially breast cancer, prostate cancer, lymphoma or  
 CC leukaemia) or a propensity to carcinoma by determination of the sequence  
 CC of a CA gene, or by determination of CA gene expression in particular  
 CC tissues. CA nucleic acids, proteins and antibodies are also useful as  
 CC therapeutic agents and in screening and evaluating drug candidates. The  
 CC present sequence represents a specifically claimed murine CA nucleic acid  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 72732 BP; 19858 A; 16168 C; 16398 G; 20003 T; 0 U; 305 Other;  
 Query Match 87.0%; Score 17.4; DB 9; Length 72732;  
 Best Local Similarity 94.7%; Pred. No. 4.3e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 AGCCAAACAGGAAACCAAC 20  
 Db 6897 AGCCAAACAGGAAACCAAC 6915  
 RESULT 8  
 ADB72425  
 ID ADB72425 standard; DNA; 72732 BP.  
 XX  
 AC ADB72425;  
 XX  
 DT 04-DEC-2003 (first entry)  
 XX  
 DE Mouse Flil gene.  
 XX  
 KW mouse; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;  
 KW cancer; neoplasm; adenocarcinoma; sarcoma; gene.  
 XX  
 OS Mus sp.  
 XX  
 PN WO2003008583-A2.  
 XX  
 PD 30-JAN-2003.  
 XX  
 PF 26-DEC-2001; 2001WO-US051291.  
 XX  
 PR 02-MAR-2001; 2001US-00798586.  
 PR 23-OCT-2001; 2001US-00004113.  
 PR 08-NOV-2001; 2001US-00052482.  
 PR 30-NOV-2001; 2001US-00997722.  
 PR 20-DEC-2001; 2001US-00034650.  
 XX  
 PA (SAGR-) SAGRES DISCOVERY.  
 XX  
 PI Morris DW, Engelhard EK;  
 XX  
 XX

DR WPI; 2003-239337/23.  
 XX  
 PT New recombinant nucleic acid, useful for treating carcinomas, lymphomas,  
 PT cancers, neoplasm, adenocarcinoma, or sarcomas.  
 XX  
 PS Claim 1; SEQ ID NO 253; 2304pp; English.  
 XX  
 CC The invention relates to a novel recombinant nucleic acid comprising a  
 CC nucleotide sequence selected from any of the 660 sequences fully defined  
 CC in the specification. A polynucleotide of the invention has cytostatic  
 CC activity, and may have a use in gene therapy, or in a vaccine. The  
 CC recombinant nucleic acids and polypeptides are useful for treating  
 CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and  
 CC sarcomas. The present sequence represents a mouse gene of the invention.  
 XX  
 SQ Sequence 72732 BP; 19858 A; 16168 C; 16398 G; 20003 T; 0 U; 305 Other;  
 Query Match 87.0%; Score 17.4; DB 10; Length 72732;  
 Best Local Similarity 94.7%; Pred. No. 4.3e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 AGCCAAACAGGAAACCAAC 20  
 Db 6897 AGCCAAACAGGAAACCAAC 6915  
 RESULT 9  
 ADE95935  
 ID ADE95935 standard; DNA; 72732 BP.  
 XX  
 AC ADE95935;  
 XX  
 DT 12-FEB-2004 (first entry)  
 XX  
 DE Mouse Flil gene genomic DNA sequence.  
 XX  
 KW cancer diagnosis; cancer treatment; carcinoma; cytostatic; gene therapy;  
 KW lymphoma; breast cancer; prostate cancer; leukaemia; ds; mouse; murine;  
 KW Flil.  
 XX  
 OS Mus sp.  
 XX  
 PN WO2003039484-A2.  
 XX  
 PD 15-MAY-2003.  
 XX  
 PF 08-NOV-2002; 2002WO-US036071.  
 XX  
 PR 08-NOV-2001; 2001US-00052482.  
 XX  
 PA (SAGR-) SAGRES DISCOVERY.  
 XX  
 PI Morris DW, Engelhard EK;  
 XX  
 WPI; 2003-441462/41.  
 XX  
 XX New carcinoma associated nucleic acids and proteins, useful for screening  
 PT drug candidates, or for diagnosing and treating carcinomas, e.g.  
 PT lymphoma, breast cancer, prostate cancer or leukemia.  
 XX  
 PS Claim 1; SEQ ID NO 193; 793pp; English.  
 XX  
 CC This invention relates to novel recombinant nucleic acids for use in  
 CC diagnosis and treatment of cancer, especially carcinomas, as well as the  
 CC use of compositions in screening methods. The compositions of the  
 CC invention may have cytostatic activity whilst the disclosed sequences may  
 CC be useful for gene therapy. The carcinoma associated nucleic acids and  
 CC proteins are useful for diagnosing and treating carcinomas, for example  
 CC lymphoma, breast cancer, prostate cancer or leukaemia, or for screening  
 CC drug candidates or bioactive agents capable of binding to, or modulating  
 CC the activity of, a carcinoma associated protein. The present sequence is  
 CC the genomic DNA sequence of the mouse Flil gene which is a carcinoma  
 CC associated gene of the invention.

```
XX SQ Sequence 72732 BP; 19858 A; 16168 C; 16398 G; 20003 T; 0 U; 305 Other;
Query Match 87.0%; Score 17.4; DB 10; Length 72732;
Best Local Similarity 94.7%; Pred. No. 4.3e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 AGCCAAACAGGACCAAC 20
|||||
Db 6897 AGCCAAACAGGACCAAC 6915
|||||

RESULT 10
ADG70447_2
Continuation (3 of 5) of ADG70447 from base 200001 (Human ANGB-CLLD8-CLLD7 hybrid gene.
WP Sequence split into 5 fragments LOCUS ADG70447 Accession Adg70447
WP Fragment Name Begin End
WP ADG70447_0 1 110000
WP ADG70447_1 100001 210000
WP ADG70447_2 200001 310000
WP ADG70447_3 300001 410000
WP ADG70447_4 400001 410846
Query Match 87.0%; Score 17.4; DB 10; Length 110000;
Best Local Similarity 94.7%; Pred. No. 4.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GAGCCAAACAGGAACCAA 19
|||||
Db 90923 GAGCAAAACAGGAACCAA 90941
|||||

RESULT 11
ABZ79565_2
Continuation (3 of 5) of ABZ79565 from base 200001 (CLLD8 and NY-REN-34 encoding DNA. )
WP Sequence split into 5 fragments LOCUS ABZ79565 Accession Abz79565
WP Fragment Name Begin End
WP ABZ79565_0 1 110000
WP ABZ79565_1 100001 210000
WP ABZ79565_2 200001 310000
WP ABZ79565_3 300001 410000
WP ABZ79565_4 400001 410846
Query Match 87.0%; Score 17.4; DB 10; Length 110000;
Best Local Similarity 94.7%; Pred. No. 4.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GAGCCAAACAGGAACCAA 19
|||||
Db 90923 GAGCAAAACAGGAACCAA 90941
|||||

RESULT 12
ADS99460/c
ID ADS99460 standard; DNA; 263852 BP.
XX
AC ADS99460;
XX
DT 30-DEC-2004 (first entry)
XX
DE Murine kinase suppressor of Ras-2 (KSR-2) genomic DNA.
XX
KW kinase suppressor of Ras-2; KSR-2; cell signalling; chromosome 5;
KW single nucleotide polymorphism; SNP; ds; gene.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT variation replace(113353,C)
FT /*tag= a
FT /standard_name= "Single nucleotide polymorphism"
XX
PN WO2004087903-A2.
```

```
XX 14-OCT-2004.
PD
XX 29-MAR-2004; 2004WO-US009487.
PF
XX 28-MAR-2003; 2003US-0457928P.
XX
PR 31-JUL-2003; 2003US-0491283P.
PR
XX (AMHP ) WYETH.
XX
XX Liu W, Wu L, Channavajhala PL, Lin L, Zhang Y;
PI
XX WPI; 2004-737702/72.
XX
DR P-PSDB; ADS99459.
DR
XX New isolated Kinase Suppressor of Ras (KSR-2) nucleic acids and
PT polypeptides, useful regulating Cot/Tp12-mediated cellular functions, or
PT polypeptides, useful regulating Cot/Tp12-mediated cellular functions, or
PT in screening assays to identify pharmacological agents for modulating KSR
PT -2 activity.
XX
XX Claim 14; SEQ ID NO 6; 475pp; English.
XX
CC The invention relates to a novel isolated kinase suppressor of Ras-2 (KSR
CC -2) nucleic acid molecule comprising SEQ ID NO. 1 or 4, a nucleic acid
CC that specifically hybridises to SEQ ID NO. 1 or 4, or its complement or a
CC protein comprising SEQ ID NO. 2 or 5. The nucleic acids and polypeptides
CC of the invention may be useful for the regulation of Cot/Tp12-mediated
CC cellular functions, including ERK and NF-kappaB activation and IL-8
CC production. The polynucleotides of the invention may be used as
CC hybridisation probes and primers to isolate nucleic acids having
CC sequences identical to or similar to those encoding the polypeptides, or
CC their allelic variants. The polynucleotides and polypeptides may be used
CC in screening assays to identify pharmacological agents or lead compounds
CC for agents capable of modulating KSR-2 activity. The current sequence is
CC that of the murine KRS-2 genomic DNA of the invention which is located at
CC chromosome 5F. Murine KRS-2 is known to contain 287 single nucleotide
CC polymorphisms; only one of these occurs in the coding region. Although
CC this sequence is described as that of a mouse, there is a significant
CC lack of homology between the encoded protein and the protein given,
CC therefore the CDS feature has been excluded.
XX
SQ Sequence 263852 BP; 64443 A; 62994 C; 63884 G; 69794 T; 0 U; 2737 Other;
Query Match 85.0%; Score 17; DB 13; Length 263852;
Best Local Similarity 100.0%; Pred. No. 7.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAGCCAAACAGGAACCA 17
|||||
Db 209107 GAGCCAAACAGGAACCA 209091
|||||

RESULT 13
ABZ40103
ID ABZ40103 standard; DNA; 483 BP.
XX
AC ABZ40103;
XX
DT 07-MAR-2003 (first entry)
XX
DE N. gonorrhoeae nucleotide sequence SEQ ID 4795.
XX
KW Antibacterial; infection; vaccine; gene therapy; gene; ds.
XX
OS Neisseria gonorrhoeae.
XX
PN WO200279243-A2.
XX
PD 10-OCT-2002.
XX
PF 12-FEB-2002; 2002WO-IB002069.
XX
XX 12-FEB-2001; 2001GB-00003424.
XX
```





KW anorectic; cardiant; hypotensive; antiinfertility; hepatotropic;  
 KW autoimmune; inflammatory; AIDS; allergy; atopic dermatitis; arthritis;  
 KW thyroiditis; bacterial infection; viral; parasitic; protozoal; fungal;  
 KW metabolic; obesity; reproductive; infertility; neurological;  
 KW Parkinson's disease; Alzheimer's; stroke; epilepsy; cardiovascular;  
 KW myocardial infarction; hypertension; eye; cell proliferative; cancer;  
 KW atherosclerosis; hepatitis; ss; gene; SNP;  
 XX single nucleotide polymorphism.  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT variation replace(285,G)  
 FT /tag= a  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT variation replace(302,G)  
 FT /tag= b  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT variation replace(2461,A)  
 FT /tag= c  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT variation replace(2484,A)  
 FT /tag= d  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT variation replace(2485,A)  
 FT /tag= e  
 FT /standard\_name= "Single nucleotide polymorphism"  
 XX WO2003094843-A2.  
 XX  
 PD 20-NOV-2003.  
 XX  
 PF 06-MAY-2003; 2003WO-US0114076.  
 XX  
 PR 10-MAY-2002; 2002US-0379840P.  
 PR 17-MAY-2002; 2002US-0381291P.  
 PR 24-MAY-2002; 2002US-0383183P.  
 PR 05-JUL-2002; 2002US-0394146P.  
 XX  
 PA (INCY-) INCYTE CORP.  
 XX  
 PI Khare R, Elliott VS, Marquis JP, Ramkumar J, Chawla NK, Mason PM;  
 PI Hafalia AJA, Swarnakar A, Jin P, Becha SD, Kable AE, Tran UK;  
 PI Baughn WR, Burford N, Gaul RC, Emerling BM, Sprague WW, Griffin JA;  
 PI Ison CH;  
 XX  
 DR WPI; 2004-011995/01.  
 DR P-PSDB; ADH18888.  
 XX  
 PT New human cell adhesion and extracellular matrix proteins and  
 PT polynucleotides, useful for diagnosing, preventing or treating diseases  
 PT or conditions associated with aberrant protein expression, e.g. cancer,  
 PT AIDS or stroke.  
 XX  
 PS Claim 5; SEQ ID NO 46; 308pp; English.  
 XX  
 CC The invention relates to a novel isolated human cell adhesion and  
 CC extracellular matrix (CADECM) polypeptide. The polypeptide of the  
 CC invention demonstrates cytostatic, antiarteriosclerotic, anti-HIV,  
 CC antiallergic, cerebroprotective, antiparkinsonian, anticonvulsant,  
 CC neurotropic, neuroprotective, antiinflammatory, ophthalmological,  
 CC antithyroid, antiarthritic, antibacterial, virucide, protozoacide,  
 CC antiparasitic, fungicide, anorectic, cardiac, hypotensive,  
 CC antiinfertility and hepatotropic activities. The polypeptide may be  
 CC useful in diagnosing, preventing or treating diseases or conditions such  
 CC as autoimmune or inflammatory disorders including AIDS, allergies, atopic  
 CC dermatitis, arthritis or thyroiditis, infections particularly bacterial,  
 CC viral, parasitic, protozoal or fungal, metabolic disorders such as  
 CC obesity, reproductive disorders e.g. infertility, neurological disorders  
 CC including Parkinson's disease, Alzheimer's disease, stroke or epilepsy,  
 CC cardiovascular disorders e.g. myocardial infarction or hypertension, eye  
 CC disorders or cell proliferative diseases including cancer,  
 CC atherosclerosis and hepatitis. The current sequence is that of the human

CC CADECM cDNA of the invention.  
 XX  
 SQ Sequence 2509 BP; 500 A; 732 C; 765 G; 512 T; 0 U; 0 Other;  
 Query Match 84.0%; Score 16.8; DB 12; Length 2509;  
 Best Local Similarity 90.0%; Pred. NO. 6.1e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 GAGCCAAACAGGAAACCAAC 20  
 |||||  
 DB 160 GAGCCAAACAGGAAACAGAC 141  
 RESULT 16  
 ADS19326/C  
 ID ADS19326 standard; cDNA; 2742 BP.  
 XX  
 AC ADS19326;  
 XX  
 DT 18-NOV-2004 (first entry)  
 XX  
 DE Human integrin beta7 cDNA sequence SeqID 35.  
 XX  
 KW human; gene; ss; alphaE integrin; CD103; alphaE-beta7;  
 KW lymphocyte migration; inflammatory bowel disease; Crohn's disease;  
 KW gastroenteritis; pulmonary inflammatory disease; asthma;  
 KW chronic bronchitis; graft rejection; psoriasis; eczema; urticaria;  
 KW scleroderma; autoimmune disease; multiple sclerosis; diabetes;  
 KW glomerulonephritis; autoimmune thyroiditis; Behcet's syndrome;  
 KW viral infection; cancer; neoplastic disease; leukaemia; lymphoma;  
 KW antiinflammatory; antiasthmatic; immunosuppressive; antipsoriatic;  
 KW dermatological; neuroprotective; antidiabetic; nephrotropic; virucide;  
 KW cytostatic; vasotropic.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US2003232387-A1.  
 XX  
 PD 18-DEC-2003.  
 XX  
 PF 14-JUN-2002; 2002US-00173551.  
 XX  
 PR 14-JUN-2002; 2002US-00173551.  
 XX  
 PA (MILL-) MILLENNIUM PHARM INC.  
 XX  
 PI Lu C;  
 XX  
 DR WPI; 2004-178738/17.  
 DR P-PSDB; ADS19327.  
 XX  
 PT Novel antibody which binds activated alpha-E integrin, specifically to  
 PT activation-induced epitope on integrin alpha-E chain (CD103), useful for  
 PT treating inflammatory bowel diseases e.g., Crohn's disease,  
 PT gastroenteritis.  
 XX  
 PS Example 1; SEQ ID NO 37; 67pp; English.  
 XX  
 CC This invention relates to novel antibody and antigen-binding fragments of  
 CC antibodies that bind alphaE-beta7 integrin chains (CD103). Specifically,  
 CC it refers to the binding of activation induced epitopes present on  
 CC activated alphaE integrins, where these integrins are activated by  
 CC exposure to divalent cations (e.g. Mn2+), to phorbol esters or suitable  
 CC growth factors and/ or mitogens. The present invention describes the  
 CC alphaE-beta7 integrin as a homing receptor that mediates lymphocyte  
 CC migration to mucosal epithelium. As such, compositions of this invention  
 CC are useful for treating a subject having an inflammatory bowel disease  
 CC such as Crohn's disease or gastroenteritis or pulmonary inflammatory  
 CC diseases such as asthma and chronic bronchitis. Furthermore, such  
 CC compositions can be used for inhibiting graft rejection, psoriasis,  
 CC eczema, urticaria, scleroderma, autoimmune diseases such as multiple  
 CC sclerosis, diabetes, glomerulonephritis, autoimmune thyroiditis, Behcet's  
 CC syndrome, viral infections, cancer and/or neoplastic diseases such as

CC leukaemias and lymphomas. Accordingly, they exhibit antiinflammatory,  
 CC antiasthmatic, immunosuppressive, antipsoriatic, dermatological,  
 CC neuroprotective, antidiabetic, nephrotropic, virucide, cytostatic and  
 CC vasotropic activities. This polynucleotide sequence is the human integrin  
 CC beta7 cDNA sequence of the invention.

XX Sequence 2742 BP; 551 A; 798 C; 833 G; 560 T; 0 U; 0 Other;  
 SQ Query Match 84.0%; Score 16.8; DB 13; Length 2742;  
 Best Local Similarity 90.0%; Pred. No. 6.1e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAAC 20  
 DB 84 GAGCCAAACAGGAACAGAC 65  
 AC  
 AC  
 ADI31813;  
 17-JUN-2004 (first entry)  
 Human cDNA #1139.

XX Human; gene; ss; immunological response; immunopathological condition;  
 KW Crohn's disease; asthma; ulcerative colitis; hypereosinophilia;  
 KW irritable bowel syndrome; osteoarthritis; rheumatoid arthritis;  
 KW acute monocytic leukaemia; antiinflammatory; antiasthmatic; antiulcer;  
 KW osteopathic; antiarthritic; antirheumatic; cytostatic.  
 XX Homo sapiens.  
 OS US6607879-B1.  
 XX 19-AUG-2003.  
 XX 09-FEB-1998; 98US-00023655.  
 XX 09-FEB-1998; 98US-00023655.  
 XX (INCY-) INCYTE CORP.  
 XX Cocks BG, Stuart SG, Seilhamer JJ;  
 XX WPI; 2003-895307/82.

XX A composition comprising a plurality of cDNAs, useful for detecting  
 PT altered expression of genes in an immunological response or for  
 PT diagnosing and treating an immunopathology, e.g. Crohn's disease, asthma  
 PT or osteoarthritis.  
 XX Claim 1; SEQ ID NO 1139; 50pp; English.  
 PS The invention relates to a composition comprising a plurality of cDNAs  
 CC for detecting the altered expression of genes in an immunological  
 CC response. The invention also relates to a method of diagnosing or  
 CC monitoring the treatment of an immunopathological condition in a sample,  
 CC comprising obtaining nucleic acids from a sample, contacting the nucleic  
 CC acids of the sample with an array comprising the plurality of cDNAs under  
 CC conditions to form one or more hybridisation complexes, detecting the  
 CC hybridisation complexes and comparing the levels of the detected  
 CC hybridisation complexes with the level of hybridisation complexes  
 CC detected in a non-diseased sample, where an altered level of the detected  
 CC hybridisation complexes correlates with the presence of an  
 CC immunopathological condition. Also disclosed are an expression profile  
 CC comprising a microarray and a plurality of detectable complexes and a  
 CC method for identifying a plurality of polynucleotide probes. The cDNAs  
 CC are useful as hybridisable array elements in a microarray for monitoring  
 CC the expression of target polynucleotides. The microarray can be used in  
 CC the diagnosis of an immunopathology, such as Crohn's disease, asthma,

CC ulcerative colitis, hypereosinophilia, irritable bowel syndrome,  
 CC osteoarthritis, rheumatoid arthritis or acute monocytic leukaemia, and in  
 CC identifying agents for the treatment of the diseases. The microarray may  
 CC also be used in drug discovery and development, toxicological and  
 CC carcinogenicity studies, forensics or pharmacogenomics. The composition  
 CC may also be used in purification of a subpopulation of mRNAs, cDNAs or  
 CC genomic fragments. This sequence represents a human cDNA of the  
 CC invention. Note: The sequence data for this patent did not form part of  
 CC the printed specification but was obtained in electronic format directly  
 CC from USPTO at seqdata.uspto.gov/sequence.html.

XX Sequence 2780 BP; 559 A; 810 C; 842 G; 569 T; 0 U; 0 Other;  
 SQ Query Match 84.0%; Score 16.8; DB 11; Length 2780;  
 Best Local Similarity 90.0%; Pred. No. 6.1e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAAC 20  
 DB 122 GAGCCAAACAGGAACAGAC 103  
 AC  
 AC  
 ADS83880;  
 11-AUG-2005 (first entry)  
 Human lymph node cDNA #1139.

XX ss; gene; human; immunological response; blood cell; cancer;  
 KW immunopathological; AIDS; allergy; anaemia; asthma; atherosclerosis;  
 KW bronchitis; ulcerative colitis; diabetes; multiple sclerosis;  
 KW osteoporosis; pancreatitis; infection; arthritis; lymph node.  
 XX Homo sapiens.  
 OS US2004077003-A1.  
 XX 22-APR-2004.  
 XX 14-AUG-2003; 2003US-00641643.  
 XX 09-FEB-1998; 98US-00023655.  
 XX (INCY-) INCYTE CORP.  
 XX Cocks BG, Stuart SG, Seilhamer JJ;  
 XX WPI; 2004-387937/36.  
 XX New compositions having a number of first, second and third  
 PT polynucleotide probes, useful in research and diagnostic applications in  
 PT cancer and immunological conditions e.g. AIDS, diabetes, osteoporosis and  
 PT infections.  
 XX Claim 15; SEQ ID NO 1139; 16pp; English.

XX The invention relates to polynucleotides which are used as probes to  
 CC detect genes differentially expressed in an immunological response,  
 CC abundantly expressed in an immunological response and/or coding for a  
 CC polypeptide known to regulate blood cell biology. The polynucleotides are  
 CC useful in research and diagnostic applications particularly in cancer and  
 CC immunopathological conditions, such as AIDS, allergies, anæmia, asthma,  
 CC atherosclerosis, bronchitis, ulcerative colitis, diabetes, multiple  
 CC sclerosis, osteoporosis, pancreatitis, infections and arthritis. The  
 CC present sequence represents a human lymph node cDNA used to detect blood  
 CC cell and immunological response gene expression. Note: The present  
 CC sequence does not appear in the printed specification but was obtained in  
 CC electronic format from the USPTO web site  
 CC (seqdata.uspto.gov/sequence.html?docID=20040077003).

```
XX SQ Sequence 2780 BP; 559 A; 810 C; 842 G; 569 T; 0 U; 0 Other;
Query Match 84.0%; Score 16.8; DB 13; Length 2780;
Best Local Similarity 90.0%; Pred. No. 6.1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGACCAAC 20
DB 122 GAGCCAAACAGGAAACAGAC 103

RESULT 19
ABL65835/c
ID ABL65835 standard; DNA; 2798 BP.
XX ABL65835;
AC ABL65835;
DT 15-MAY-2002 (first entry)
XX
DE Lung cancer related gene sequence SEQ ID NO:4172.
XX
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200194629-A2.
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US010838.
XX
PR 05-JUN-2000; 2000US-0209473P.
PR 05-JUN-2000; 2000US-0209531P.
PR 18-SEP-2000; 2000US-0233133P.
PR 18-SEP-2000; 2000US-0233617P.
PR 20-SEP-2000; 2000US-0234009P.
PR 20-SEP-2000; 2000US-0234034P.
PR 20-SEP-2000; 2000US-0234052P.
PR 22-SEP-2000; 2000US-0234509P.
PR 22-SEP-2000; 2000US-0234567P.
PR 25-SEP-2000; 2000US-0234923P.
PR 25-SEP-2000; 2000US-0234924P.
PR 25-SEP-2000; 2000US-0235077P.
PR 25-SEP-2000; 2000US-0235082P.
PR 25-SEP-2000; 2000US-0235134P.
PR 25-SEP-2000; 2000US-0235280P.
PR 26-SEP-2000; 2000US-0235637P.
PR 26-SEP-2000; 2000US-0235638P.
PR 27-SEP-2000; 2000US-0235711P.
PR 27-SEP-2000; 2000US-0235720P.
PR 27-SEP-2000; 2000US-0235840P.
PR 27-SEP-2000; 2000US-0235863P.
PR 28-SEP-2000; 2000US-0236028P.
PR 28-SEP-2000; 2000US-0236032P.
PR 28-SEP-2000; 2000US-0236033P.
PR 28-SEP-2000; 2000US-0236034P.
PR 28-SEP-2000; 2000US-0236109P.
PR 28-SEP-2000; 2000US-0236111P.
PR 29-SEP-2000; 2000US-0236842P.
PR 29-SEP-2000; 2000US-0236891P.
PR 02-OCT-2000; 2000US-0237172P.
PR 02-OCT-2000; 2000US-0237173P.
PR 02-OCT-2000; 2000US-0237278P.
PR 02-OCT-2000; 2000US-0237294P.
PR 02-OCT-2000; 2000US-0237295P.
PR 03-OCT-2000; 2000US-0237316P.
PR 03-OCT-2000; 2000US-0237425P.
PR 03-OCT-2000; 2000US-0237598P.
PR 03-OCT-2000; 2000US-0237604P.

PR 03-OCT-2000; 2000US-0237606P.
PR 03-OCT-2000; 2000US-0237608P.
PR 01-NOV-2000; 2000US-0244867P.
PR 01-NOV-2000; 2000US-0245084P.
XX (AVAL-) AVALON PHARM.
XX
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX WPI; 2002-188264/24.
XX
PT Screening for anti-neoplastic agent involves exposing cells to a chemical
PT agent to be tested for anti-neoplastic activity, and determining a change
PT in expression of a gene of a signature gene set.
XX
PS Claim 1; SEQ ID NO 4172; 44pp; English.
XX
CC The present invention describes a method (M1) for screening for an anti-
CC neoplastic agent. The method involves exposing cells to a chemical agent
CC to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening an
CC anti-neoplastic agent, and can be used for producing a product which is
CC the data collected with respect to the anti-neoplastic agent as a result
CC of M1, and the data is sufficient to convey the chemical structure and/or
CC properties of the agent. M1 can be used in the treatment of cancer such
CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
CC tumour
XX
SQ Sequence 2798 BP; 577 A; 810 C; 842 G; 569 T; 0 U; 0 Other;
Query Match 84.0%; Score 16.8; DB 6; Length 2798;
Best Local Similarity 90.0%; Pred. No. 6.1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGACCAAC 20
DB 122 GAGCCAAACAGGAAACAGAC 103

RESULT 20
ADS99921/c
ID ADS99921 standard; cDNA; 2798 BP.
XX ADS99921;
AC ADS99921;
DT 02-DEC-2004 (first entry)
XX
DE Human integrin, beta 7 (ITGB7), cDNA.
XX
KW Human; ss; gene; Bisulphite; metastasis; cancer; cytostatic;
KW DNA methylation; matrix-assisted laser desorption/ionisation; MALDI;
KW electrospray; mass spectrometry; CpG dinucleotide; solid tumour.
XX
OS Homo sapiens.
XX
PN US2003148327-A1.
XX
PD 07-AUG-2003.
XX
PF 21-JAN-2003; 2003US-00240485.
XX
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
```

PR 06-APR-2001; 2001WO-EP003970.  
 XX (OLEK/) OLEK A.  
 PA (PIEP/) PIEPENBROCK C.  
 PA (BERL/) BERLIN K.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 XX WPI; 2002-010922/01.  
 DR GENBANK; NM\_000889.  
 XX  
 PT New nucleic acid derived from chemically treated metastasis genes, useful  
 PT for diagnosis of cancers by analysis of cytosine methylation, also for  
 PT treatment.  
 XX  
 XX Claim 2; Page; 9pp; English.  
 XX  
 CC The invention relates to a nucleic acid comprising at least 18 bases from  
 CC a segment of the chemically pretreated DNA of genes associated with  
 CC metastasis, i.e. any of ADS99709-ADS99906 human genomic sequences or any  
 CC of the 19 sequences appearing as ADS99911-ADS99929. SEQ ID 2,4,6 etc are  
 CC the complements of SEQ ID 1,3,5, etc. Also included are an oligomer  
 CC (particularly an oligonucleotide or peptide nucleic acid) comprising at  
 CC least one base sequence of at least 9 bases which hybridises to (or is  
 CC identical with) the sequences referred to above, producing an array of  
 CC the oligomers on a carrier, obtaining genetic and/or epigenetic  
 CC parameters for diagnosis and/or therapy of diseases (or predisposition to  
 CC them) by analysis of cytosine methylation and a kit comprising a  
 CC bisulphite (disulphite or hydrogen sulphite) and the oligomers. In the  
 CC method of above 5-methylated cytosines in a genomic DNA sample are  
 CC converted chemically to uracil, or another base with hybridisation  
 CC properties different from those of cytosine, then fragments of the  
 CC treated DNA amplified (particularly by polymerase chain reaction) using  
 CC the oligomers and a polymerase (preferably heat stable) to produce  
 CC labelled amplicons. These are tested for hybridisation to an array of  
 CC oligomers and any hybridisation detected. The amplicons are labelled with  
 CC fluorescent or radioactive markers, or with a detachable mass marker to  
 CC allow their detection by mass spectrometry, specifically using the matrix  
 CC assisted laser desorption/ionisation (MALDI) or electrospray techniques.  
 CC To improve detection in the mass spectrometer, fragments formed in the  
 CC instrument have only a single net charge (positive or negative). The  
 CC genomic DNA is from e.g. a cell line, biopsy sample, blood, or paraffin-  
 CC embedded tissue sample. Oligonucleotides or peptide-nucleic acids that  
 CC are complementary to (or identical with) parts of the nucleic acids listed  
 CC above may be used as primers for amplification of the nucleic acids or  
 CC their complements, and for determining cytosine methylation status and/or  
 CC single nucleotide polymorphisms in metastasis-related genes. They can be  
 CC used for analysis of diseases associated with methylation of CpG  
 CC dinucleotides and to determine (epi)genetic parameters for diagnosis  
 CC and/or therapy of disease (or predisposition). The genomic DNA sequences  
 CC are useful for diagnosis and therapy of solid tumours and cancer. The  
 CC present sequence is human gene fragment or cDNA associated with  
 CC metastasis. Note: The sequence was not displayed in the specification but  
 CC was obtained by the Indexer from GenBank.  
 XX  
 SQ Sequence 2798 BP; 577 A; 810 C; 842 G; 569 T; 0 U; 0 Other;  
 Query Match 84.0%; Score 16.8; DB 7; Length 2798;  
 Best Local Similarity 90.0%; Pred. No. 6.1e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 GAGCCAAACAGGAACCAAC 20  
 |||||  
 DB 122 GAGCCAAACAGGAACAGAC 103  
 |||||  
 RESULT 21  
 ABX70638/c  
 ID ABX70638 standard; cDNA; 2798 BP.  
 XX  
 AC ABX70638;  
 XX  
 DT 04-MAR-2003 (first entry)

XX Human cDNA encoding NOV4a.  
 DE  
 XX  
 KW Human; ss; gene; NOVX; metabolic disorder; diabetes; cardiomyopathy;  
 KW obesity; infectious disease; anorexia; neurodegenerative disorder;  
 KW Alzheimer's disease; Parkinson's disease; immune disorder;  
 KW hematopoietic disorder; dyslipidaemia; metabolic disturbance;  
 KW metabolic syndrome X; wasting disorder; cancer; gene therapy; SNP;  
 KW single nucleotide polymorphism.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT variation replace(423,A)  
 FT /\*tag= a  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT variation replace(1133,T)  
 FT /\*tag= b  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT variation replace(1261,A)  
 FT /\*tag= c  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT variation replace(2126,G)  
 FT /\*tag= d  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT variation replace(2241,C)  
 FT /\*tag= e  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT variation replace(2285,T)  
 FT /\*tag= f  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT variation replace(2303,A)  
 FT /\*tag= g  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT variation replace(2414,T)  
 FT /\*tag= h  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT variation replace(2443,A)  
 FT /\*tag= i  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT variation replace(2447,T)  
 FT /\*tag= j  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT  
 FT WO200281518-A2.  
 PN 17-OCT-2002.  
 PD  
 XX 21-FEB-2002; 2002WO-US005374.  
 PF  
 XX 21-FEB-2001; 2001US-0270220P.  
 XX 21-FEB-2001; 2001US-0270523P.  
 PR 23-FEB-2001; 2001US-0270797P.  
 PR 23-FEB-2001; 2001US-0270810P.  
 PR 08-MAR-2001; 2001US-0274295P.  
 PR 16-MAR-2001; 2001US-0276400P.  
 PR 16-MAR-2001; 2001US-0276677P.  
 PR 26-MAR-2001; 2001US-0278796P.  
 PR 04-APR-2001; 2001US-0281521P.  
 PR 25-APR-2001; 2001US-0286548P.  
 PR 17-MAY-2001; 2001US-0291765P.  
 PR 10-AUG-2001; 2001US-0311595P.  
 PR 13-AUG-2001; 2001US-0311980P.  
 PR 10-SEP-2001; 2001US-0318526P.  
 PR 17-SEP-2001; 2001US-0322712P.  
 PR 18-OCT-2001; 2001US-0330307P.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 XX Pena CEA, Shinkets RA, Li L, Shenoy SG, Kekuda R, Spytek KA;  
 PI Vernet CAM, Malyankar UL, Guo X, Gusev VY, Casman SJ, Boldog FI;  
 PI Furtak X, Tchernev VI, Patturajan M, Gangolli EA, Pattigar M, Liu X;  
 PI Baumgartner JC, Gerlach VL, Spaderna SK, Zerhusen BD;

XX WPI; 2003-046859/04.  
 DR P-PSDB; ABU52583.  
 XX  
 XX New isolated NOVX polypeptide useful for treating cardiomyopathy,  
 PT atherosclerosis, metabolic disorders, diabetes, obesity, infectious  
 PT disease, anorexia, neurodegenerative disorders, Alzheimer's disease and  
 PT cancer.  
 XX  
 XX Claim 9; Page 34; 479pp; English.  
 XX  
 XX The invention relates to an isolated polypeptide termed NOVX (NOV1, 2a,  
 CC 2b, 3a, 3b, 4a, 4b, 5, 6, 7a-c, 8a-e, 9a-b, 10, 11, 12a-c, 13, 14, 15,  
 CC 16a-d, 17a-b, 18, 19, 20a-b, 21-30) appearing as ABU52578-ABU52624), a  
 CC variant of NOVX, a mature form of NOVX, and a variant of the mature form  
 CC of NOVX. Also included are a nucleic acid molecule (NOVX NA) encoding  
 CC NOVX, or a fragment or complement of NOVX NA, a vector comprising NOVX  
 CC NA, a cell comprising the vector, an anti-NOVX antibody (ab), determining  
 CC the presence or amount of NOVX or NOVX NA in a sample, and identifying an  
 CC agent that binds or modulates the expression or activity of NOVX. NOVX,  
 CC NOVX NA or ab is useful for treating or preventing a NOVX-associated  
 CC disorder in a subject, preferably human. Ab is useful for determining the  
 CC presence or amount of NOVX in a sample. NOVX is useful for identifying an  
 CC agent that binds to NOVX. NOVX, NOVX NA or ab is useful for treating  
 CC metabolic disorders, diabetes, cardiomyopathy, obesity, infectious  
 CC disease, anorexia, neurodegenerative disorders, Alzheimer's disease,  
 CC Parkinson's disease, immune disorders, haematopoietic disorders, and  
 CC various dyslipidaemias, metabolic disturbances associated with obesity,  
 CC the metabolic syndrome X and wasting disorders associated with chronic  
 CC diseases, various cancers, endocrine, connective tissue, blood, vascular,  
 CC skin, renal, bone, brain, muscle disorders, or bacterial, fungal,  
 CC protozoal or viral infections. NOVX, NOVX NA or ab is useful in screening  
 CC assays, detection assays, predictive medicine, and in methods of  
 CC treatment. NOVX is useful as immunogen, to screen for potential  
 CC ant/agonist compounds, and as bait protein in a two-hybrid or three-  
 CC hybrid assay. NOVX NA is useful in gene therapy, to express NOVX, to  
 CC detect NOVX mRNA or a genetic lesion in a NOVX gene, and to modulate NOVX  
 CC activity. The cell is useful for producing non-human transgenic animals.  
 CC Ab is useful for isolating, and purifying NOVX and to monitor protein  
 CC levels in tissue as part of a clinical testing procedure. The present  
 CC sequence encodes a NOVX protein  
 XX  
 XX Sequence 2798 BP; 577 A; 810 C; 842 G; 569 T; 0 U; 0 Other;  
 SQ  
 Query Match 84.0%; Score 16.8; DB 10; Length 2798;  
 Best Local Similarity 90.0%; Pred. No. 6.1e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 GAGCCAAACAGGAACCAAC 20  
 Db ||||| ||||| ||||| |||||  
 122 GAGCCAAACAGGAACAGAC 103  
 RESULT 22  
 ADP10372/C  
 ID ADP10372 standard; DNA; 2798 BP.  
 XX  
 XX AC ADP10372;  
 XX  
 XX 05-MAY-2005 (first entry)  
 XX  
 XX DE DNA encoding a PRO polypeptide, SEQ ID NO 5343.  
 XX  
 XX Antinflammatory; Immune disorder; Dermatological; Immunosuppressive;  
 KW Antirheumatic; Antiarthritic; Osteopathic; Hemostatic; Antianemic;  
 KW Antithyroid; Antidiabetic; Nephrotropic; CNS-Gen.; Hepatotropic;  
 KW Virucide; Gastrointestinal-Gen.; Antipsoriatic; Antiasthmatic;  
 KW Antiallergic; ds; gene; diagnosis.  
 XX  
 XX OS Homo sapiens.  
 XX  
 XX PN WO2005016962-A2.  
 XX  
 XX PD 24-FEB-2005.  
 XX  
 XX PF 11-AUG-2004; 2004WO-US026249.  
 XX  
 XX PR 11-AUG-2003; 2003US-0493546P.  
 XX  
 XX PA (GETH ) GENENTECH INC.  
 XX  
 XX PI Abbas A, Clark H, Ouyang W, Williams MP, Wood WI, Wu TD;  
 XX

XX 24-APR-2002; 2002US-00131831.  
 PR 20-DEC-2002; 2002US-00325899.  
 XX  
 XX (EXPR-) EXPRESSION DIAGNOSTICS INC.  
 XX Wohlgenuth J, Fry K, Woodward R, Ly N, Prentice J, Morris M;  
 PI Rosenberg S;  
 XX  
 XX WPI; 2004-400724/37.  
 DR  
 XX  
 XX Diagnosing or monitoring transplant rejection, e.g. heart, kidney, liver,  
 PT pancreas, pancreatic islet, lung, bone marrow or stem cell transplant  
 PT rejection, in an individual, comprises detecting the expression level of  
 PT the genes.  
 XX  
 XX Claim 80; SEQ ID NO 381; 1762pp; English.  
 XX  
 XX The present invention relates to diagnosing or monitoring transplant  
 CC rejection, e.g. cardiac or kidney transplant rejection, in an individual  
 CC comprises detecting the expression level of one or more genes. The  
 CC methods, system and kits are useful in diagnosing or monitoring  
 CC transplant rejection, e.g. heart, kidney, liver, pancreas, pancreatic  
 CC islet, lung, bone marrow or stem cell transplant rejection,  
 CC xenotransplant rejection or mechanical organ replacement rejection, in an  
 CC individual. The method is also useful in assessing the immune status of  
 CC an individual. The methods are also useful in diagnosing and monitoring  
 CC diseases that involve the immune system, e.g. rheumatoid arthritis,  
 CC lupus, inflammatory bowel diseases, multiple sclerosis, HIV/AIDS or  
 CC viral, bacterial or fungal infection. The present sequence represents a  
 CC mRNA reference sequence for a 50 mer oligonucleotide marker for diagnosis  
 CC and monitoring of allograft rejection and other disorders.  
 XX  
 XX Sequence 2798 BP; 577 A; 810 C; 842 G; 569 T; 0 U; 0 Other;  
 SQ  
 Query Match 84.0%; Score 16.8; DB 12; Length 2798;  
 Best Local Similarity 90.0%; Pred. No. 6.1e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 GAGCCAAACAGGAACCAAC 20  
 Db ||||| ||||| ||||| |||||  
 122 GAGCCAAACAGGAACAGAC 103  
 RESULT 23  
 ADP19537/C  
 ID ADP19537 standard; DNA; 2798 BP.  
 XX  
 XX AC ADP19537;  
 XX  
 XX 05-MAY-2005 (first entry)  
 XX  
 XX DE DNA encoding a PRO polypeptide, SEQ ID NO 5343.  
 XX  
 XX Antinflammatory; Immune disorder; Dermatological; Immunosuppressive;  
 KW Antirheumatic; Antiarthritic; Osteopathic; Hemostatic; Antianemic;  
 KW Antithyroid; Antidiabetic; Nephrotropic; CNS-Gen.; Hepatotropic;  
 KW Virucide; Gastrointestinal-Gen.; Antipsoriatic; Antiasthmatic;  
 KW Antiallergic; ds; gene; diagnosis.  
 XX  
 XX OS Homo sapiens.  
 XX  
 XX PN WO2005016962-A2.  
 XX  
 XX PD 24-FEB-2005.  
 XX  
 XX PF 11-AUG-2004; 2004WO-US026249.  
 XX  
 XX PR 11-AUG-2003; 2003US-0493546P.  
 XX  
 XX PA (GETH ) GENENTECH INC.  
 XX  
 XX PI Abbas A, Clark H, Ouyang W, Williams MP, Wood WI, Wu TD;  
 XX

XX WPI; 2005-182330/19.  
XX  
PT New nucleic acid encoding PRO polypeptide, useful for diagnosing and  
PT treating an immune related disorder, e.g. systemic lupus erythematosus,  
PT rheumatoid arthritis, osteoarthritis, thyroiditis, or diabetes mellitus.  
XX  
PS Claim 1; SEQ ID NO 5343; 158pp; English.  
XX  
CC The invention relates to an isolated nucleic acid encoding a PRO  
CC polypeptide. The polypeptide, agonist or an antagonist, antibody,  
CC composition, and method are useful for diagnosing and treating an immune  
CC related disorder, e.g. systemic lupus erythematosus, rheumatoid  
CC arthritis. The present sequence represents a PRO polypeptide.  
XX  
SQ Sequence 2798 BP; 577 A; 810 C; 842 G; 569 T; 0 U; 0 Other;

Query Match 84.0%; Score 16.8; DB 14; Length 2798;  
Best Local Similarity 90.0%; Pred. No. 6.1e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAAC 20  
Db 122 GAGCCAAACAGGAACAGAC 103

RESULT 24  
ADV15613/c  
ID ADV15613 standard; DNA; 2798 BP.  
XX  
AC ADV15613;  
XX  
DT 05-MAY-2005 (first entry)  
XX  
DE DNA encoding a PRO polypeptide, SEQ ID NO 1419.  
XX  
KW Antiinflammatory; Immune disorder; Dermatological; Immunosuppressive;  
KW Antirheumatic; Antiarthritic; Osteopathic; Hemostatic; Antianemic;  
KW Antithyroid; Antidiabetic; Nephrotropic; CNS-Gen.; Hepatotrophic;  
KW Virucide; Gastrointestinal-Gen.; Antipsoriatic; Antiasthmatic;  
KW Antiallergic; ds; gene; diagnosis.  
XX  
OS Homo sapiens.  
XX  
PN WO2005016962-A2.  
XX  
PD 24-FEB-2005.  
XX  
PF 11-AUG-2004; 2004WO-US026249.  
XX  
PR 11-AUG-2003; 2003US-0493546P.  
XX  
PA (GETH ) GENENTECH INC.  
XX  
PI Abbas A, Clark H, Ouyang W, Williams MP, Wood WI, Wu TD;  
XX  
DR WPI; 2005-182330/19.  
XX  
PT New nucleic acid encoding PRO polypeptide, useful for diagnosing and  
PT treating an immune related disorder, e.g. systemic lupus erythematosus,  
PT rheumatoid arthritis, osteoarthritis, thyroiditis, or diabetes mellitus.  
XX  
PS Claim 1; SEQ ID NO 1419; 158pp; English.  
XX  
CC The invention relates to an isolated nucleic acid encoding a PRO  
CC polypeptide. The polypeptide, agonist or an antagonist, antibody,  
CC composition, and method are useful for diagnosing and treating an immune  
CC related disorder, e.g. systemic lupus erythematosus, rheumatoid  
CC arthritis. The present sequence represents a DNA encoding a PRO  
CC polypeptide.  
XX  
SQ Sequence 2798 BP; 577 A; 810 C; 842 G; 569 T; 0 U; 0 Other;

Query Match 84.0%; Score 16.8; DB 14; Length 2798;  
Best Local Similarity 90.0%; Pred. No. 6.1e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAAC 20  
Db 122 GAGCCAAACAGGAACAGAC 103

RESULT 25  
ACN43200/c  
ID ACN43200 standard; cDNA; 2800 BP.  
XX  
AC ACN43200;  
XX  
DT 18-NOV-2004 (first entry)  
XX  
DE Human diagnostic and therapeutic polynucleotide SEQ ID NO:2075.  
XX  
KW ss; gene; gene therapy; human diagnostic and therapeutic polynucleotide;  
KW dithp.  
XX  
OS Homo sapiens.  
XX  
PN WO2004023973-A2.  
XX  
PD 25-MAR-2004.  
XX  
PF 12-SEP-2003; 2003WO-US028227.  
XX  
PR 12-SEP-2002; 2002US-0410259P.  
PR 12-SEP-2002; 2002US-0410260P.  
XX  
PA (INCY-) INCYTE CORP.  
XX  
PI Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;  
PI Harthshorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV;  
PI Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;  
PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstin EH;  
PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;  
PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton ES;  
PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;  
PI Patury S, Shi X, Suarez CJ;  
XX  
DR WPI; 2004-329368/30.  
DR P-PSDB; ABM84548.  
XX  
PT New diagnostic and therapeutic polynucleotides and polypeptides, useful  
PT in diagnosing a condition, disease or disorder associated with human  
PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or  
PT in gene mapping.  
XX  
PS Claim 1; Page; 190pp; English.  
XX  
CC The invention relates to novel diagnostic and therapeutic polynucleotides  
CC selected from one of the 2722 sequences defined in the specification. A  
CC polynucleotide of the invention may have a use in gene therapy. The human  
CC polynucleotide and therapeutic polynucleotides (dithp) or polypeptides may be  
CC used to diagnose a particular condition, disease or disorder associated  
CC with human molecules, e.g. cell proliferative disorders,  
CC autoimmune/inflammatory disorder, developmental disorder,  
CC disorder, neurological disorders, gastrointestinal disorders, or  
CC infections caused by virus, bacteria, fungi or parasite. The dithp  
CC molecules may also be used in genetic mapping, in identifying individuals  
CC from minute biological samples, in detecting single nucleotide  
CC polymorphisms, as molecular weight markers, and for somatic or germline  
CC gene therapy. The present sequence represents a dithp polynucleotide of  
CC the invention. Note: The sequence data for this patent is not represented  
CC in the printed specification, but was obtained in electronic format  
CC directly from WIPO at [www.wipo.int/pct/en/sequences/listing.htm](http://www.wipo.int/pct/en/sequences/listing.htm)  
XX  
SQ Sequence 2800 BP; 556 A; 815 C; 827 G; 602 T; 0 U; 0 Other;



```
Query Match      84.0%; Score 16.8; DB 13; Length 2800;
Best Local Similarity 90.0%; Pred. No. 6.1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAAC 20
   |||||
Db 97 GAGCCAAACAGGAACAGAC 78

RESULT 26
AAS94871/c
ID AAS94871 standard; DNA; 2835 BP.
XX
AC AAS94871;
XX
DT 14-FEB-2002 (first entry)
XX
DE Human DNA sequence #126 expressed during foam cell differentiation.
XX
KW Human; foam cell differentiation; atherosclerosis; cerebral stroke;
KW cardiovascular disorder; coronary artery disease; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200177389-A2.
XX
PD 18-OCT-2001.
XX
PF 04-APR-2001; 2001WO-US011128.
XX
PR 05-APR-2000; 2000US-0195106P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Shiffman D, Somogyi R, Lawn R, Seilhamer JJ, Porter GJ, Mikita T;
PI Tai J;
XX
WPI; 2002-010925/01.
XX
Composition useful for diagnosis of conditions, disorders or diseases
PT associated with atherosclerosis, comprises several polynucleotides that
PT are differentially expressed in foam cell development.
XX
Claim 1; Page 183; 315pp; English.
XX
The present invention relates to the isolation of human polynucleotide
CC sequences that are differentially expressed during foam cell
CC differentiation. The polynucleotide sequences of the invention or a
CC composition comprising these polynucleotides are useful as a high
CC throughput method for detecting altered expression of one or more
CC polynucleotides in a sample. The polynucleotides can be used in the
CC diagnosis of disorders associated with foam cell development such as
CC atherosclerosis, cerebral stroke, and cardiovascular disorders such as
CC coronary artery disease. The polynucleotide sequences can also be used as
CC PCR primers and probes. The polynucleotides of the invention are also
CC useful in gene therapy. AAS94746-AAS95021 represent the human
CC polynucleotide sequences of the invention which are differentially
CC expressed during foam cell differentiation
XX
Sequence 2835 BP; 582 A; 826 C; 853 G; 574 T; 0 U; 0 Other;

Query Match      84.0%; Score 16.8; DB 6; Length 2835;
Best Local Similarity 90.0%; Pred. No. 6.1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAAC 20
   |||||
Db 162 GAGCCAAACAGGAACAGAC 143

RESULT 27
ACN43199/c
ID ACN43199 standard; cDNA; 3244 BP.

Query Match      84.0%; Score 16.8; DB 13; Length 3244;
Best Local Similarity 90.0%; Pred. No. 6.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAAC 20
   |||||
Db 97 GAGCCAAACAGGAACAGAC 78

RESULT 28
AAD22781/c
ID AAD22781 standard; DNA; 31766 BP.

XX
AC ACN43199;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human diagnostic and therapeutic polynucleotide SEQ ID NO:2074.
XX
KW ss; gene; gene therapy; human diagnostic and therapeutic polynucleotide;
KW dithp.
XX
OS Homo sapiens.
XX
PN WO2004023973-A2.
XX
PD 25-MAR-2004.
XX
PF 12-SEP-2003; 2003WO-US028227.
XX
PR 12-SEP-2002; 2002US-0410259P.
XX
PA (INCY-) INCYTE CORP.
XX
PI Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
PI Harthshorne TA, Suchorolski MT, Altus CM, Pitts SJ, Eldor LV;
PI Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstin EH;
PI Peralta CH, Anderson SB, Rioux P, Shen EU, Wu MC, Stuve LL;
PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton ES;
PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI Patury S, Shi X, Suarez CJ;
XX
WPI: 2004-329368/30.
DR P-PSDB; AEM84547.
XX
New diagnostic and therapeutic polynucleotides and polypeptides, useful
PT in diagnosing a condition, disease or disorder associated with human
PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
PT in gene mapping.
XX
Claim 1; Page; 190pp; English.
XX
The invention relates to novel diagnostic and therapeutic polynucleotides
CC selected from one of the 2722 sequences defined in the specification. A
CC polynucleotide of the invention may have a use in gene therapy. The human
CC polynucleotide and therapeutic polynucleotides (dithp) or polypeptides may be
CC used to diagnose a particular condition, disease or disorder associated
CC with human molecules, e.g. cell proliferative disorders,
CC autoimmune/inflammatory disorder, developmental disorder, endocrine
CC disorder, neurological disorders, gastrointestinal disorders, or
CC infections caused by virus, bacteria, fungi or parasite. The dithp
CC molecules may also be used in genetic mapping, in identifying individuals
CC from minute biological samples, in detecting single nucleotide
CC polymorphisms, as molecular weight markers, and for somatic or germline
CC gene therapy. The present sequence represents a dithp polynucleotide of
CC the invention. Note: The sequence data for this patent is not represented
CC in the printed specification, but was obtained in electronic format
CC directly from WIPO at www.wipo.int/pct/en/sequences/listing.htm
XX
Sequence 3244 BP; 641 A; 937 C; 977 G; 689 T; 0 U; 0 Other;
```



XX AAD22781;  
AC  
XX DT 26-FEB-2002 (first entry)  
XX DE Human sulphate transporter protein splice form genomic DNA.  
XX  
XX KW Human; sulphate transporter protein splice form; immunostimulant;  
KW gene therapy; pharmacogenomic analysis; drug screening; diagnosis;  
KW tissue typing; gene expression; ds.  
XX OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT exon 914..1147  
FT /\*tag= a  
FT /number= 1  
FT intron 1148..2008  
FT /\*tag= b  
FT /number= 1  
FT exon 2009..2286  
FT /\*tag= c  
FT /number= 2  
FT intron 2287..2517  
FT /\*tag= d  
FT /number= 2  
FT exon 2518..2563  
FT /\*tag= e  
FT /number= 3  
FT intron 2564..5066  
FT /\*tag= f  
FT /number= 3  
FT exon 5067..5146  
FT /\*tag= g  
FT /number= 4  
FT intron 5147..7047  
FT /\*tag= h  
FT /number= 4  
FT exon 7048..7190  
FT /\*tag= i  
FT /number= 5  
FT intron 7191..115250  
FT /\*tag= j  
FT /number= 5  
FT exon 15251..15426  
FT /\*tag= k  
FT /number= 6  
FT intron 15427..15844  
FT /\*tag= l  
FT /number= 6  
FT exon 15845..15917  
FT /\*tag= m  
FT /number= 7  
FT intron 15918..18499  
FT /\*tag= n  
FT /number= 7  
FT exon 18500..18550  
FT /\*tag= o  
FT /number= 8  
FT intron 18551..20881  
FT /\*tag= p  
FT /number= 8  
FT exon 20882..20942  
FT /\*tag= q  
FT /number= 9  
FT intron 20943..21839  
FT /\*tag= r  
FT /number= 9  
FT exon 21840..21895  
FT /\*tag= s  
FT /number= 10  
FT intron 21896..22194  
FT /\*tag= t

FT exon /number= 10  
FT 22195..22335  
FT /\*tag= u  
FT /number= 11  
FT intron 22336..23816  
FT /\*tag= v  
FT /number= 11  
FT exon 23817..23944  
FT /\*tag= w  
FT /number= 12  
FT intron 23945..24261  
FT /\*tag= x  
FT /number= 12  
FT exon 24262..24361  
FT /\*tag= y  
FT /number= 13  
FT intron 24362..24840  
FT /\*tag= z  
FT /number= 13  
FT exon 24841..24974  
FT /\*tag= aa  
FT /number= 14  
FT intron 24975..27018  
FT /\*tag= ab  
FT /number= 14  
FT exon 27019..27091  
FT /\*tag= ac  
FT /number= 15  
FT intron 27092..28484  
FT /\*tag= ad  
FT /number= 15  
FT exon 28485..28576  
FT /\*tag= ae  
FT /number= 16  
XX  
PN WO200181412-A2.  
XX  
XX PD 01-NOV-2001.  
XX  
XX PF 26-APR-2001; 2001WO-US013380.  
XX  
XX PR 26-APR-2000; 2000US-0199811P.  
XX PR 02-AUG-2000; 2000US-00630719.  
XX PR 22-JAN-2001; 2001US-00765344.  
XX  
XX PA (PEXE ) PE CORP NY.  
XX  
XX PI Wei M, Ketchum K, Di Francesco V, Beasley EM;  
XX WPI; 2002-026157/03.  
XX  
XX PT New transporter proteins related to sulfate transporter subfamily and  
XX nucleic acid molecule encoding the protein for diagnosing, treating  
XX disease or condition mediated by transporter protein and identifying  
XX modulators.  
XX  
XX PS Claim 4; Fig 3; 90pp; English.  
XX  
XX CC The invention relates to human sulphate transporter protein splice forms  
XX and their corresponding cDNA molecules. Human transporter protein and its  
XX cDNA are used as models and as targets for the development of human  
XX therapeutic agents, in the identification of therapeutic proteins that  
XX modulate transporter activity in cells and tissues that express the  
XX transporter. Transporter protein and its DNA are also useful in  
XX competition binding assays to discover compounds that interact with the  
XX transporter. The transporter-modulating agents are useful in an animal or  
XX other model to determine the efficacy, toxicity, mechanism of action or  
XX side effects of treatment with such an agent. The transporter proteins  
XX also provide a target for diagnosing a disease or predisposition to a  
XX disease mediated by the peptide. Human sulphate transporter proteins in  
XX cells or tissues are used to determine the pattern of expression of the  
XX protein among various tissues in an organism, to assess abnormal tissue  
XX distribution or abnormal expression during development or progression of

CC a biological condition, to assess normal and aberrant subcellular  
CC localisation of cells in various tissues in an organism, in  
CC pharmacogenomic analysis, for tissue typing and for inhibiting protein  
CC function. Human transporter DNA is useful for drug screening to identify  
CC compounds that modulate transporter nucleic acid expression. The present  
CC sequence is human sulphate transporter protein splice form genomic DNA  
XX  
SQ Sequence 31766 BP; 6622 A; 8135 C; 8441 G; 8318 T; 0 U; 250 Other;  
Query Match 84.0%; Score 16.8; DB 6; Length 31766;  
Best Local Similarity 90.0%; Pred. No. 7.6e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GAGCCAAACAGGAAACCAAC 20  
Db 11047 GAGCAACAGGAAACCAAC 11028  
RESULT 29  
AAL50687/C  
ID AAL50687 standard; DNA; 31766 BP.  
XX  
AC AAL50687;  
XX  
DT 16-JAN-2003 (first entry)  
XX  
DE Human sulphate transporter protein genomic DNA sequence.  
XX  
KW Human; gene; ds; splice form; sulphate transporter protein;  
KW transporter-related condition; germ cell tumour; brain oligodendrogloma;  
KW single nucleotide polymorphism; SNP.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 914..28576  
FT FT /\*tag= a  
FT FT /product= "Human sulphate transporter protein"  
FT FT 914..1147  
FT FT /\*tag= b  
FT FT /number= 1  
FT FT 1148..2008  
FT FT /\*tag= c  
FT FT /number= 1  
FT FT 2009..2286  
FT FT /\*tag= d  
FT FT /number= 2  
FT FT 2287..2517  
FT FT /\*tag= e  
FT FT /number= 2  
FT FT 2518..2563  
FT FT /\*tag= f  
FT FT /number= 3  
FT FT 2564..5066  
FT FT /\*tag= g  
FT FT /number= 3  
FT FT replace(4181, G)  
FT FT /\*tag= h  
FT FT /note= "Single nucleotide polymorphism"  
FT FT replace(4519, T)  
FT FT /\*tag= i  
FT FT /note= "Single nucleotide polymorphism"  
FT FT 5067..5146  
FT FT /\*tag= j  
FT FT /number= 4  
FT FT 5147..7047  
FT FT /\*tag= k  
FT FT /number= 4  
FT FT 7048..7190  
FT FT /\*tag= l  
FT FT /number= 5  
FT FT 7191..15250  
FT FT /\*tag= m

FT variation /number= 5  
FT replace(10009, C)  
FT /\*tag= n  
FT /note= "Single nucleotide polymorphism"  
FT replace(12072, C)  
FT /\*tag= o  
FT /note= "Single nucleotide polymorphism"  
FT replace(12210, C)  
FT /\*tag= p  
FT /note= "Single nucleotide polymorphism"  
FT replace(13376, T)  
FT /\*tag= q  
FT /note= "Single nucleotide polymorphism"  
FT 15251..15426  
FT /\*tag= r  
FT /number= 6  
FT 15427..15844  
FT /\*tag= s  
FT /number= 6  
FT 15845..15917  
FT /\*tag= t  
FT /number= 7  
FT 15918..18499  
FT /\*tag= u  
FT /number= 7  
FT replace(16256, T)  
FT /\*tag= v  
FT /note= "Single nucleotide polymorphism"  
FT 18500..18550  
FT /\*tag= w  
FT /number= 8  
FT 18551..20881  
FT /\*tag= x  
FT /number= 8  
FT 20882..20942  
FT /\*tag= y  
FT /number= 9  
FT 20943..21839  
FT /\*tag= z  
FT /number= 9  
FT replace(20952, C)  
FT /\*tag= aa  
FT /note= "Single nucleotide polymorphism"  
FT replace(20987, C)  
FT /\*tag= ab  
FT /note= "Single nucleotide polymorphism"  
FT replace(21620, G)  
FT /\*tag= ac  
FT /note= "Single nucleotide polymorphism"  
FT replace(21795, C)  
FT /\*tag= ad  
FT /note= "Single nucleotide polymorphism"  
FT 21840..21895  
FT /\*tag= ae  
FT /number= 10  
FT 21896..22194  
FT /\*tag= af  
FT /number= 10  
FT 22195..22335  
FT /\*tag= ag  
FT /number= 11  
FT 22336..23816  
FT /\*tag= ah  
FT /number= 11  
FT replace(22753, T)  
FT /\*tag= ai  
FT /note= "Single nucleotide polymorphism"  
FT replace(22945, A)  
FT /\*tag= aj  
FT /note= "Single nucleotide polymorphism"  
FT replace(23032, A)  
FT /\*tag= ak  
FT /note= "Single nucleotide polymorphism"  
FT

FT variation replace(23738, A)  
FT /\*tag= al  
FT /note= "Single nucleotide polymorphism"  
FT exon 23817. .23944  
FT /\*tag= am  
FT /number= 12  
FT intron 23945. .24261  
FT /\*tag= an  
FT /number= 12  
FT variation replace(23952, G)  
FT /\*tag= ao  
FT /note= "Single nucleotide polymorphism"  
FT variation replace(24123, A)  
FT /\*tag= ap  
FT /note= "Single nucleotide polymorphism"  
FT exon 24262. .24361  
FT /\*tag= aq  
FT /number= 13  
FT intron 24362. .24840  
FT /\*tag= ar  
FT /number= 13  
FT variation replace(24691, T)  
FT /\*tag= as  
FT /note= "Single nucleotide polymorphism"  
FT exon 24841. .24974  
FT /\*tag= at  
FT /number= 14  
FT intron 24975. .27018  
FT /\*tag= au  
FT /number= 14  
FT variation replace(25015, A)  
FT /\*tag= av  
FT /note= "Single nucleotide polymorphism"  
FT variation replace(25191, T)  
FT /\*tag= aw  
FT /note= "Single nucleotide polymorphism"  
FT exon 27019. .27091  
FT /\*tag= ax  
FT /number= 15  
FT intron 27092. .28484  
FT /\*tag= ay  
FT /number= 15  
FT exon 28485. .28576  
FT /\*tag= az  
FT /number= 16  
FT WO200259306-A2.  
FT 01-AUG-2002.  
FT 26-OCT-2001; 2001WO-US042809.  
FT 22-JAN-2001; 2001US-00765344.  
FT (APPL-) APPLERA CORP.  
FT Wei M, Ketchum KA, Di Francesco V, Beasley EM;  
FT WPI; 2002-599785/64.  
FT New isolated human transporter proteins, useful for developing  
FT therapeutic or diagnostic compositions, particularly for developing human  
FT therapeutic agents that modulate transporter activity in cells or  
FT tissues, or treat tumors.  
FT Claim 4; Page 66-75; 108pp; English.  
FT  
FT The invention comprises the amino acid and coding sequences of two splice  
FT forms of a novel human sulphate transporter protein. The DNA and protein  
FT sequences of the invention are useful for treating transporter-related  
FT conditions (e.g. germ cell tumours or brain oligodendrogliomas). The  
FT present DNA sequence represents the genomic region comprising the gene  
FT encoding the human sulphate transporter protein

XX  
XX Query Match 84.0%; Score 16.8; DB 6; Length 31766;  
XX Best Local Similarity 90.0%; Pred. No. 7.6e+02;  
XX Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
XX  
XX QY 1 GAGCCAAACAGGAAACCAAC 20  
XX ||| ||||| ||||| |||||  
XX Db 11047 GAGACAAACAGGAAACCAAC 11028  
XX  
XX RESULT 30  
XX ADJ35017  
XX ID ADJ35017 standard; DNA; 1071 BP.  
XX AC ADJ35017;  
XX XX  
XX DT 22-APR-2004 (first entry)  
XX DE DNA encoding xylanase from an environmental sample seq id 233.  
XX KW antibacterial; fungicide; thermostable xylanase activity;  
XX KW dough conditioning; beverage production; nutritional supplement;  
XX KW animal feed; lignin reduction; wood product; xylan; bacterial infection;  
XX KW fungal infection; coccidiosis; gene; ds.  
XX OS Unidentified.  
XX XX  
XX PN WO2003106654-A2.  
XX PD 24-DEC-2003.  
XX XX  
XX PF 16-JUN-2003; 2003WO-US019153.  
XX PR 14-JUN-2002; 2002US-0389299P.  
XX PA (DIVE-) DIVERSA CORP.  
XX PI Steer B, Callen W, Healey S, Hazlewood G, Wu D, Blum D;  
XX PI Esteghlalian A;  
XX XX  
XX DR WPI; 2004-099016/10.  
XX DR P-PSDB; ADJ35018.  
XX PT Novel xylanase recombinant polypeptide useful for improving textile  
XX texture, treating paper, eliminating microorganisms.  
XX PS Claim 1; SEQ ID NO 233; 570pp; English.  
XX CC The invention describes an isolated or recombinant polypeptide (I),  
XX CC having 50% or more identity to 190 300-1200 residue amino acid sequences  
XX CC (S1), given in the specification, over a region of 100 or more residues  
XX CC and the polypeptide as thermostable xylanase activity. (I) is useful for:  
XX CC dough conditioning; beverage production; as a nutritional supplement in  
XX CC animal feed; reducing lignin in a wood or a wood product; and for  
XX CC eliminating and protecting animals from a microorganism comprising xylan.  
XX CC The polynucleotide (II) encoding (I) is useful for amplifying nucleic  
XX CC acid encoding a polypeptide having a xylanase activity which involves  
XX CC amplification of a template nucleic acid with a primer pair capable of  
XX CC amplifying (II) or its subsequence. (I) is useful for treating and  
XX CC preventing bacterial infection and fungal infection e.g. coccidiosis.  
XX CC This sequence encodes xylanase protein isolated from an environmental  
XX CC sample.  
XX SQ Sequence 1071 BP; 237 A; 299 C; 342 G; 193 T; 0 U; 0 Other;  
XX  
XX Query Match 82.0%; Score 16.4; DB 12; Length 1071;  
XX Best Local Similarity 94.4%; Pred. No. 8.6e+02;  
XX Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
XX  
XX QY 1 GAGCCAAACAGGAAACCAAC 18  
XX ||| ||||| ||||| |||||  
XX Db 93 GAGCCAGACAGGAAACCAAC 110

```

RESULT 31
ADA70575/c
ID ADA70575 standard; DNA; 1794 BP.
XX AC ADA70575;
XX AC ADA70575;
DT 20-NOV-2003 (first entry)
XX DE Rice gene, SEQ ID 3898.
XX KW Plant; bacterial infection; fungal infection; viral infection; rice;
XX KW Gene; ds.
XX OS Oryza sativa.
XX PN WO2003000898-A1.
XX PD 03-JAN-2003.
XX PF 22-JUN-2001; 2001WO-IB001105.
XX PR 22-JUN-2001; 2001WO-IB001105.
XX PA (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
XX PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX DR WPI; 2003-175290/17.
XX PT Identifying at least one gene involved in plant resistance or response to
XX PT pathogenic infection for conferring resistance or tolerance to a plant to
XX PT bacterial, fungal or viral infection by determining or detecting plant
XX PT gene expression.
XX PS Claim 6; SEQ ID NO 3898; 899pp; English.
XX CC The present invention relates to a method (M1) for identifying genes
XX CC involved in plant resistance or response to pathogenic infection. M1
XX CC comprises identifying a gene whose expression is significantly altered in
XX CC the incompatible interaction of plant gene expression relative to
XX CC expression of the gene in an uninfected plant, in a mutant plant that
XX CC does not express a gene associated with response to pathogenic infection,
XX CC or in a corresponding incompatible or compatible interaction. (M1) is
XX CC useful for conferring resistance to resistance or tolerance to a plant to
XX CC bacterial, fungal or viral infection. The present sequence was used to
XX CC illustrate the invention.
XX SQ Sequence 1794 BP; 533 A; 371 C; 460 G; 429 T; 0 U; 1 Other;

Query Match 82.0%; Score 16.4; DB 8; Length 1794;
Best Local Similarity 94.4%; Pred. No. 9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GCCAAACAGAACCAAAAC 20
Db 143 GCCAAACAGAACCAAAAC 126

RESULT 32
ACL28714/c
ID ACL28714 standard; cDNA; 1794 BP.
XX AC ACL28714;
XX AC ACL28714;
DT 02-JUN-2005 (first entry)
XX DE Rice abiotic stress responsive polynucleotide SEQ ID NO:2670.
XX KW as; abiotic stress tolerance; transgenic plant; plant; cereal;
XX KW agriculture.

```

```

XX Oryza sativa.
XX OS WO2003008540-A2.
XX PN 30-JAN-2003.
XX PD 21-JUN-2002; 2002WO-US019668.
XX PF 22-JUN-2001; 2001US-0300112P.
XX PR 24-AUG-2001; 2001US-0314662P.
XX PR 26-SEP-2001; 2001US-0325277P.
XX PR 21-NOV-2001; 2001US-0332132P.
XX PA (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX PI Krepis J, Briggs SP, Cooper B, Glazebrook J, Goff SA, Katagiri F;
XX PI Moughamer T, Provart N, Rieke D, Zhu T;
XX DR WPI; 2003-248011/24.
XX PF New stress-responsive nucleic acid, useful for altering the
XX PT responsiveness of a plant, e.g. cereal, to an abiotic stress such as cold
XX PT stress, salt stress or osmotic stress.
XX PR Claim 1; SEQ ID NO 2670; 89pp; English.
XX CC The invention relates to novel abiotic stress responsive polynucleotides
XX CC and polypeptides. Also disclosed are vectors, expression cassettes, host
XX CC cells, and plants containing such polynucleotides. Also disclosed are
XX CC methods for using the polynucleotides and polypeptides to alter the
XX CC responsiveness of a plant to abiotic stress. The invention is useful in
XX CC agriculture. The nucleic acid is useful for determining whether a test
XX CC plant has been exposed to an abiotic stress condition. It is also useful
XX CC for selecting an agent that alters abiotic stress regulated
XX CC polynucleotide expression in a plant cell, and to identify a homolog or
XX CC ortholog to an abiotic stress responsive polynucleotide. The nucleic acid
XX CC molecule and the polypeptide encoded by it are useful in altering the
XX CC responsiveness of a plant to an abiotic stress, such as cold stress, salt
XX CC stress, osmotic stress or any of their combinations. The present sequence
XX CC is used in the exemplification of the invention
XX SQ Sequence 1794 BP; 533 A; 371 C; 460 G; 429 T; 0 U; 1 Other;

Query Match 82.0%; Score 16.4; DB 11; Length 1794;
Best Local Similarity 94.4%; Pred. No. 9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GCCAAACAGAACCAAAAC 20
Db 143 GCCAAACAGAACCAAAAC 126

RESULT 33
ADW10224
ID ADW10224 standard; DNA; 2476 BP.
XX AC ADW10224;
XX AC ADW10224;
DT 24-MAR-2005 (first entry)
XX DE Colon proliferative disorder associated human DNA fragment, SEQ ID 29.
XX KW colorectal tumor; CpG methylation detection; cytostatic; gene therapy;
XX KW proliferative disorder; carcinoma; ds.
XX OS Homo sapiens.
XX PN US2004265833-A1.
XX PD 30-DEC-2004.
XX PR 23-JUN-2003; 2003US-00602494.

```

```
XX PR 23-JUN-2003; 2003US-00602494.
XX PA (LOFT/) LOFTON-DAY C.
XX PA (SLED/) SLEDZIEWSKI A.
XX PA (THOM/) THOMAS J.
XX PA (DAYR/) DAY R W.
XX PA (TONN/) TONNES-PRIDDY L.
XX PA (CARD/) CARDON K.
XX PI Lofton-Day C, Sledziewski A, Thomas J, Day RW, Tonnes-Priddy L;
XX PI Cardon K;
XX WPI; 2005-089566/10.
XX
XX PT Detecting and distinguishing colorectal cell proliferative disorders by
XX PT contacting genomic DNA of biological sample with reagent that
XX PT distinguishes methylated and non-methylated CpG dinucleotides within
XX PT target sequence of genomic DNA.
XX
XX PS Claim 25; SEQ ID NO 29; 23pp; English.
XX
XX CC The invention relates to a novel method for detecting and distinguishing
XX CC between, or among, colorectal cell proliferative disorders. The method
XX CC involves contacting genomic DNA of a biological sample obtained from the
XX CC subject with one or more reagent(s), or a series of reagents that
XX CC distinguishes between methylated and non-methylated CpG dinucleotides
XX CC within a target sequence of the genomic DNA. The invention further
XX CC comprises: a nucleic acid comprising a sequence of 18 or more contiguous
XX CC nucleotides of a treated genomic DNA sequence chosen from any one of 284
XX CC fully defined nucleotide sequences, whose sequence listing is not
XX CC provided in the specification, and their complementary sequences, where
XX CC the contiguous sequence has one or more CpG, TpA, or CpA dinucleotide,
XX CC and the treatment is suitable to convert one or more of the unmethylated
XX CC cytosine base(s) of the genomic DNA sequence initially to uracil or
XX CC another base that is detectably dissimilar to cytosine in terms of
XX CC hybridization; an oligomer or peptide nucleic acid (PNA)-oligomer,
XX CC comprising 9 or more contiguous nucleotides that is complementary to or
XX CC hybridizes under moderately stringent or stringent conditions to one of
XX CC the 284 DNA sequences and their complementary sequences provided in the
XX CC source document, which is treated; a set of oligomers comprising two or
XX CC more of the oligomer of PNA-oligomer; an array of oligomers; and a kit
XX CC for carrying out the above methods. The method and its novel compositions
XX CC have cytostatic activity. The polynucleotide sequence may be used in gene
XX CC therapy. The above methods are useful for detecting and distinguishing
XX CC between, or among, colorectal cell proliferative disorders chosen from
XX CC colorectal carcinoma, colon adenomas and colon polyps, in a biological
XX CC sample, such as histological slides, biopsies, paraffin embedded tissue,
XX CC bodily fluids, stool, blood, serum, plasma and their combinations. The
XX CC oligomer array is useful as a probe for detecting one or more of the
XX CC cytosine methylation state, or single nucleotide polymorphisms within the
XX CC genomic DNA or their complementary sequences. The polynucleotides of the
XX CC invention are useful for classifying, distinguishing between, or among,
XX CC diagnosing or determining the predisposition for colorectal cell
XX CC proliferative disorders, or for the therapy of colorectal cell
XX CC proliferative disorders. This polynucleotide sequence represents a DNA
XX CC fragment associated with the development of colon cell proliferative
XX CC disorders, used in the novel methylated/non-methylated CpG dinucleotide
XX CC detection method of the invention. Note: This sequence is not shown in
XX CC the specification. It has been electronically downloaded from the USPTO
XX CC website.
XX
XX SQ Sequence 2476 BP; 493 A; 787 C; 765 G; 431 T; 0 U; 0 Other;
XX
XX Query Match 82.0%; Score 16.4; DB 14; Length 2476;
XX Best Local Similarity 94.4%; Pred. No. 9.3e+02;
XX Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 3 GCCAACAGGAGCAACCAAC 20
XX |||||||||||||||
XX Db 65 GCCAACAGGAGCAACCAAC 82
XX
XX RESULT 35
XX ABD32548/c
XX ID ABD32548 standard; DNA; 310268 BP.
XX AC ABD32548;
XX XX
XX DT 18-NOV-2004 (first entry)
XX XX
XX DE Human cancer-associated genomic DNA HD14-033.
XX
XX KW Human; ds; cancer-associated protein; gene; cytostatic; cancer;
XX KW leukaemia; lymphoma; CAP.
XX OS Homo sapiens.
XX PN WO2004074320-A2.
XX XX
XX PD 02-SEP-2004.
XX XX
XX PF 17-FEB-2004; 2004WO-US004730.
XX XX
XX
XX RESULT 34
XX ADQ97298/c
XX ID ADQ97298 standard; DNA; 95484 BP.
XX AC ADQ97298;
XX XX
XX DT 07-OCT-2004 (first entry)
XX XX
XX DE Mouse cancer associated sequence MD08-028, SEQ ID 275.
XX KW Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Mouse; ds.
XX XX
XX OS Mus musculus.
XX PN WO2004060304-A2.
XX PD 22-JUL-2004.
XX XX
XX PF 22-DEC-2003; 2003WO-US041389.
XX XX
XX PR 27-DEC-2002; 2002US-00330773.
XX XX
XX PS (SAGR-) SAGRES DISCOVERY INC.
XX PI Morris DW, Malandro MS;
XX XX
XX WPI; 2004-543781/52.
XX
XX PT New isolated cancer associated nucleic acids comprising at least 10
XX PT contiguous nucleotides, useful for diagnosing, preventing and/or treating
XX PT cancers such as leukemia and lymphoma.
XX
XX PS Claim 1; SEQ ID NO 275; 199pp; English.
XX
XX CC The present invention relates to cancer associated sequences (ADQ97025-
XX CC ADQ98004). The sequences are useful for the diagnosis, prevention and/or
XX CC treatment of cancer, such as leukemia and lymphoma. Note: The sequence
XX CC data for this patent did not form part of the printed specification, but
XX CC was obtained in electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 95484 BP; 22182 A; 22693 C; 24695 G; 24779 T; 0 U; 1135 Other;
XX
XX Query Match 82.0%; Score 16.4; DB 12; Length 95484;
XX Best Local Similarity 94.4%; Pred. No. 1.3e+03;
XX Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 2 AGCCAAACAGGAGCAACCA 19
XX |||||||||||||||
XX Db 80942 AGCCAAACAGGAGCAACCA 80925
XX
XX RESULT 35
XX ABD32548/c
XX ID ABD32548 standard; DNA; 310268 BP.
XX AC ABD32548;
XX XX
XX DT 18-NOV-2004 (first entry)
XX XX
XX DE Human cancer-associated genomic DNA HD14-033.
XX
XX KW Human; ds; cancer-associated protein; gene; cytostatic; cancer;
XX KW leukaemia; lymphoma; CAP.
XX OS Homo sapiens.
XX PN WO2004074320-A2.
XX XX
XX PD 02-SEP-2004.
XX XX
XX PF 17-FEB-2004; 2004WO-US004730.
XX XX
```

PR 14-FEB-2003; 2003US-00367094.  
PR 14-MAR-2003; 2003US-0038838.  
PR 15-APR-2003; 2003US-00417375.  
PR 13-JUN-2003; 2003US-00461862.  
PR 15-SEP-2003; 2003US-00663431.  
PR 15-DEC-2003; 2003US-00737318.  
XX  
XX (SAGR-) SAGRES DISCOVERY INC.  
XX  
XX Morris DW, Morris DW, Malandro MS;  
XX WPI; 2004-652914/63.  
XX  
XX New isolated cancer-associated polynucleotides and polypeptides useful  
PT for diagnosing, preventing or treating cancers, especially lymphoma and  
PT leukemia, or in screening for agents that modulate cancer.  
XX  
XX claim 16; seqid 24; 310pp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising at least 10  
CC contiguous nucleotides of any of the 233 polynucleotide sequences given  
CC in the specification, or its complement. The nucleic acids encode cancer-  
CC associated proteins. Also included are an expression vector comprising  
CC the isolated nucleic acid cited above, a host cell comprising the above  
CC recombinant nucleic acid or expression vector, a microarray for detecting  
CC a cancer-associated (CA) nucleic acid comprising at least one probe  
CC comprising at least 10 contiguous nucleotides of any of the above-  
CC mentioned nucleotide sequences, an isolated polypeptide (encoded within  
CC an open reading frame of a CA sequence selected from any of the 95  
CC polynucleotide sequences as mentioned in the specification, or its  
CC complement), an isolated antibody, (or its antigen binding fragment) that  
CC binds to the above polypeptide, a hybridoma that produces the above  
CC monoclonal antibody, a pharmaceutical composition comprising the above  
CC antibody and a pharmaceutical excipient, a kit for detecting cancer  
CC cells (comprising the antibody cited above, methods for diagnosing cancer  
CC or for detecting the presence or absence of cancer cells in an  
CC individual, a method for inhibiting growth of cancer cells in an  
CC individual, a method for delivering a therapeutic agent to cancer cells  
CC in an individual, an electronic library comprising the above  
CC polynucleotide or polypeptide (or their fragments), methods of screening  
CC for anticancer activity or for a bioactive agent capable of modulating  
CC the activity of a CA protein (CAP), methods for detecting cancer  
CC associated with expression of a polypeptide in a test cell sample, a  
CC method for treating cancers and a method for inhibiting the expression of  
CC CA gene in a cell. The composition and methods are useful for detecting,  
CC diagnosing, preventing and treating cancers, especially lymphoma and  
CC leukaemia. These may also be used in screening for agents that modulate  
CC cancer. The present sequence is a human CAP genomic sequence. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 310268 BP; 87522 A; 60932 C; 62901 G; 98913 T; 0 U; 0 Other;  
Query Match 82.0%; Score 16.4; DB 13; Length 310268;  
Best Local Similarity 94.4%; Pred. No. 1.4e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 3 GCCAACACGAGCAACCAAC 20  
Db 11729 GCCAACACGAGCAACCAAC 11712  
RESULT 36  
AAC50252  
ID AAC50252 standard; DNA; 871 BP.  
XX  
XX AAC50252;  
AC  
XX 18-OCT-2000 (first entry)  
DT  
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 64140.  
DE  
XX

KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway; metabolic pathway;  
XX promoter; termination sequence; ss.  
OS Arabidopsis thaliana.  
XX  
XX EP1033405-A2.  
PN  
XX  
XX 06-SEP-2000.  
PD  
XX  
XX 25-FEB-2000; 2000EP-00301439.  
PF  
XX  
XX 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0123788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 28-APR-1999; 99US-0130891P.  
PR 30-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.

Hybridisation assay; genetic mapping; gene expression control;  
protein identification; signal transduction pathway; metabolic pathway;  
promoter; termination sequence; ss.

Arabidopsis thaliana.

EP1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-00301439.

25-FEB-1999; 99US-0121825P.

05-MAR-1999; 99US-0123180P.

09-MAR-1999; 99US-0123548P.

23-MAR-1999; 99US-0123788P.

25-MAR-1999; 99US-0126264P.

29-MAR-1999; 99US-0126785P.

01-APR-1999; 99US-0127462P.

06-APR-1999; 99US-0128234P.

08-APR-1999; 99US-0128714P.

16-APR-1999; 99US-0129845P.

19-APR-1999; 99US-0130077P.

21-APR-1999; 99US-0130449P.

23-APR-1999; 99US-0130510P.

28-APR-1999; 99US-0130891P.

30-APR-1999; 99US-0131449P.

30-APR-1999; 99US-0132048P.

30-APR-1999; 99US-0132407P.

04-MAY-1999; 99US-0132484P.

05-MAY-1999; 99US-0132485P.

06-MAY-1999; 99US-0132486P.

06-MAY-1999; 99US-0132487P.

07-MAY-1999; 99US-0132863P.

11-MAY-1999; 99US-0134256P.

14-MAY-1999; 99US-0134218P.

14-MAY-1999; 99US-0134219P.

14-MAY-1999; 99US-0134221P.

14-MAY-1999; 99US-0134370P.

18-MAY-1999; 99US-0134768P.

19-MAY-1999; 99US-0134941P.

20-MAY-1999; 99US-0135124P.

21-MAY-1999; 99US-0135353P.

24-MAY-1999; 99US-0135629P.

25-MAY-1999; 99US-0136021P.

27-MAY-1999; 99US-0136392P.

28-MAY-1999; 99US-0136782P.

01-JUN-1999; 99US-0137222P.

03-JUN-1999; 99US-0137528P.

04-JUN-1999; 99US-0137502P.

07-JUN-1999; 99US-0137724P.

08-JUN-1999; 99US-0138094P.

10-JUN-1999; 99US-0138540P.

10-JUN-1999; 99US-0138847P.

14-JUN-1999; 99US-0139119P.

16-JUN-1999; 99US-0139452P.

16-JUN-1999; 99US-0139453P.

17-JUN-1999; 99US-0139492P.

18-JUN-1999; 99US-0139454P.

18-JUN-1999; 99US-0139455P.

18-JUN-1999; 99US-0139456P.

18-JUN-1999; 99US-0139457P.

18-JUN-1999; 99US-0139458P.

18-JUN-1999; 99US-0139459P.

18-JUN-1999; 99US-0139460P.

18-JUN-1999; 99US-0139461P.

18-JUN-1999; 99US-0139462P.

18-JUN-1999; 99US-0139463P.

18-JUN-1999; 99US-0139750P.

18-JUN-1999; 99US-0139763P.

21-JUN-1999; 99US-0139817P.

22-JUN-1999; 99US-0139899P.

23-JUN-1999; 99US-0140353P.

PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-01411287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142877P.  
PR 13-JUL-1999; 99US-0143342P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147933P.  
PR 10-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.

PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156598P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157533P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161992P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 80.0%; Score 16; DB 3; Length 871;  
Best Local Similarity 100.0%; Pred. No. 1.3e+03;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 CCAAAACGAGAACCAA 19  
Db 829 CCAAAACGAGAACCAA 844

RESULT 37  
ABL23613  
ID ABL23613 standard; DNA; 5735 BP.  
XX  
AC ABL23613;

XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 22312.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ds.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX

PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US009231.  
XX  
PR 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX WPI; 2001-656860/75.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX  
PS Claim 1; SEQ ID NO 22312; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 5735 BP; 1645 A; 1526 C; 1463 G; 1101 T; 0 U; 0 Other;  
XX  
Query Match 80.0%; Score 16; DB 4; Length 5735;  
Best Local Similarity 100.0%; Pred. No. 1.5e+03;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 4 CCAACACGAGAACCAAA 19  
Db 1936 CCAACACGAGAACCAAA 1951  
RESULT 38  
ABL23612  
ID ABL23612 standard; DNA; 8954 BP.  
XX  
AC ABL23612;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 22309.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ds.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US009231.  
XX  
PR 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX WPI; 2001-656860/75.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell

PT interactions.  
XX  
PS Claim 1; SEQ ID NO 22309; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 8954 BP; 2717 A; 2061 C; 2015 G; 2161 T; 0 U; 0 Other;  
XX  
Query Match 80.0%; Score 16; DB 4; Length 8954;  
Best Local Similarity 100.0%; Pred. No. 1.6e+03;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 4 CCAACACGAGAACCAAA 19  
Db 3413 CCAACACGAGAACCAAA 3428  
RESULT 39  
AEB39175\_01  
Continuation (2 of 35) of AEB39175 from base 100001 (L. pneumophila DNA SEQ ID NO 3507.  
WP Sequence split into 35 fragments LOCUS AEB39175 Accession Aeb39175  
WP Fragment Name Begin End  
WP AEB39175\_00 1 110000  
WP AEB39175\_01 100001 210000  
WP AEB39175\_02 200001 310000  
WP AEB39175\_03 300001 410000  
WP AEB39175\_04 400001 510000  
WP AEB39175\_05 500001 610000  
WP AEB39175\_06 600001 710000  
WP AEB39175\_07 700001 810000  
WP AEB39175\_08 800001 910000  
WP AEB39175\_09 900001 1010000  
WP AEB39175\_10 1000001 1110000  
WP AEB39175\_11 1100001 1210000  
WP AEB39175\_12 1200001 1310000  
WP AEB39175\_13 1300001 1410000  
WP AEB39175\_14 1400001 1510000  
WP AEB39175\_15 1500001 1610000  
WP AEB39175\_16 1600001 1710000  
WP AEB39175\_17 1700001 1810000  
WP AEB39175\_18 1800001 1910000  
WP AEB39175\_19 1900001 2010000  
WP AEB39175\_20 2000001 2110000  
WP AEB39175\_21 2100001 2210000  
WP AEB39175\_22 2200001 2310000  
WP AEB39175\_23 2300001 2410000  
WP AEB39175\_24 2400001 2510000  
WP AEB39175\_25 2500001 2610000  
WP AEB39175\_26 2600001 2710000  
WP AEB39175\_27 2700001 2810000  
WP AEB39175\_28 2800001 2910000  
WP AEB39175\_29 2900001 3010000  
WP AEB39175\_30 3000001 3110000  
WP AEB39175\_31 3100001 3210000  
WP AEB39175\_32 3200001 3310000  
WP AEB39175\_33 3300001 3410000  
WP AEB39175\_34 3400001 3503610  
Query Match 80.0%; Score 16; DB 14; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 2e+03;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 4 CCAACACGAGAACCAAA 19  
|||||



Db 86502 CCAAACAGGAACCAA 86517

# RESULT 40

AEBS3722

ID AEB35722 standard; DNA; 184368 BP.

XX AC AEB35722;

XX DT 08-SEP-2005 (first entry)

XX DE L. pneumophila DNA SEQ ID NO 54.

XX KW detection; infection; Antibacterial; Vaccine; ds; gene.

XX OS Legionella pneumophila.

XX PN WO2005049642-A2.

XX PD 02-JUN-2005.

XX PF 23-SEP-2004; 2004WO-IB003578.

XX PR 21-NOV-2003; 2003PR-00013687.

XX PA (INSP ) INST PASTEUR.

XX PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.

XX PA (UPLY-) UNIV LYON 1 BERNARD CLAUDE.

XX PA (CNRS ) CNRS CENT NAT RECH SCI.

XX PI Buchrieser C, Tichit M, Etienne J, Ma L, Cazalet C, Glaser P;

XX PI Rusniok C, Bouchier C, Zidane N, Magnier A, Kunst F, Vandenesch F;

XX PI Jarraud S;

XX DR WPI; 2005-388305/40.

XX PT New genome of Legionella pneumophila Paris strain and derived

XX PT polypeptides, useful for detection or identification of the strain and

XX PT for treatment and prevention of infections.

XX PS Claim 1; SEQ ID NO 54; 660pp; English.

XX CC The invention relates to an isolated or purified nucleotide sequences (I)

CC from Legionella pneumophila Paris strain. (I), and their related

CC sequences or fragments, are useful as primers and probes for detection

CC and amplification, including differentiation between the Paris and

CC Philadelphia strains of Legionella pneumophila and to prepare recombinant

CC (hybrid) polypeptides (II). (II) are also useful for preparation of

CC specific antibodies (Ab), also used for detection/identification of

CC Legionella, and some (I), specifically those involved in synthesis of

CC surface proteins, are targets for identification of inhibitors. (II), or

CC vectors that contain (I), are useful as vaccines and immunogenic

CC compositions, for treatment and prevention of infections by L.

CC pneumophila. The present sequence represents a L. pneumophila DNA.

XX SQ Sequence 184368 BP; 57341 A; 33604 C; 36303 G; 57120 T; 0 U; 0 Other;

Query Match 80.0%; Score 16; DB 14; Length 184368;

Best Local Similarity 100.0%; Pred. No. 2e+03; 0; Indels 0; Gaps 0;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 CCAAACAGGAACCAA 19

|||||

Db 4382 CCAAACAGGAACCAA 4397

Search completed: November 20, 2005, 17:43:48

Job time : 328.938 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2005, 16:52:01 ; Search time 2991.11 Seconds  
(without alignments)  
312.841 Million cell updates/sec

Title: US-10-627-757-22

Perfect score: 20  
Sequence: 1 gagccaaacagggaacaaac 20

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_hic:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_est7:\*  
9: gb\_gsal:\*  
10: gb\_gsal2:\*  
11: gb\_gsal3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	18.4	92.0	362	10 AG203792	AG203792 Pan trogl
C 2	18.4	92.0	752	10 C2806891	C2806891 OC_Ba018
C 3	18.4	92.0	938	2 BE906007	BE906007 601497010
C 4	17.4	87.0	300	2 BB564726	BB564726 BB564726
C 5	17.4	87.0	337	3 BQ329066	BQ329066 CW3-EN004
C 6	17.4	87.0	357	9 BH288892	BH288892 CH230-2L1
C 7	17.4	87.0	386	1 AW080511	AW080511 xc24a11.x
C 8	17.4	87.0	413	10 CG363310	CG363310 CGWKD52TH
C 9	17.4	87.0	414	5 BH262971	BH262971 CH230-121
C 10	17.4	87.0	418	5 BY692431	BY692431 BY692431
C 11	17.4	87.0	430	1 AI672491	AI672491 wa03h01.x
C 12	17.4	87.0	463	1 AA607599	AA607599 vo48e05.r
C 13	17.4	87.0	477	6 CB161822	CB161822 K-EST0221
C 14	17.4	87.0	491	1 AJ679230	AJ679230 AJ679230
C 15	17.4	87.0	492	2 BE707223	BE707223 MR0-HT044
C 16	17.4	87.0	545	9 BZ634760	BZ634760 PUA6A68TD
C 17	17.4	87.0	549	6 CB157333	CB157333 K-EST0216
C 18	17.4	87.0	549	6 CB164700	CB164700 K-EST0225
C 19	17.4	87.0	553	10 CG845087	CG845087 PUA6A68TR
C 20	17.4	87.0	554	9 BZ634756	BZ634756 PUA6A68TB
C 21	17.4	87.0	554	10 CG845160	CG845160 PUA6A68TF
C 22	17.4	87.0	585	10 CW378022	CW378022 fsbb001fo

C 23	17.4	87.0	588	1 AU019801	AU019801 AU019801
C 24	17.4	87.0	628	2 BB631472	BB631472 BB631472
C 25	17.4	87.0	636	10 CL156011	CL156011 104_342_1
C 26	17.4	87.0	660	9 BZ220315	BZ220315 CH230-361
C 27	17.4	87.0	666	10 CG852549	CG852549 ZMMBBB034
C 28	17.4	87.0	666	11 CR887634	CR887634 Sub_scrof
C 29	17.4	87.0	735	9 BZ418761	BZ418761 if27a04.b
C 30	17.4	87.0	741	10 CNS01RON	AL164192 Tetraodon
C 31	17.4	87.0	777	9 BZ971033	BZ971033 PUDFL35TD
C 32	17.4	87.0	781	9 CC750285	CC750285 ZMMBBB013
C 33	17.4	87.0	791	8 DR474762	DR474762 WS00962.C
C 34	17.4	87.0	908	2 BF791575	BF791575 602251677
C 35	17.4	87.0	914	10 CNS020RK	AL175961 Tetraodon
C 36	17.4	87.0	933	11 CNS03MM7	AL250936 Tetraodon
C 37	17.4	87.0	1012	10 CG859874	CG859874 ZMMBBB026
C 38	17.4	87.0	1016	11 CNS05FQU	AL335343 Tetraodon
C 39	17.4	87.0	1022	10 CW923704	CW923704 EDCB010TR
C 40	17.4	87.0	1119	10 CW935554	CW935554 EDCCU52TR
C 41	17.4	87.0	1266	10 AG100948	AG100948 Pan trogl
C 42	17	85.0	442	10 CL374721	CL374721 RPC144_47
C 43	17	85.0	450	1 AI044560	AI044560 UT-R-C1-3
C 44	17	85.0	517	5 BU758993	BU758993 UT-R-FF0-
C 45	17	85.0	612	6 CB548361	CB548361 AMGNNUC:U

ALIGNMENTS

RESULT 1  
AG203792/c  
LOCUS  
DEFINITION Pan troglodytes DNA, clone: RP43-088K24.TJ, genomic survey  
ACCESSION AG203792  
VERSION AG203792.1 GI:45235967  
KEYWORDS GSS.  
SOURCE Pan troglodytes (chimpanzee)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Pan.  
REFERENCE  
1  
Park,H., Kim,Y., Kim,S., Han,Y., Woo,T., Park,K., Eun,C.J.,  
Hoon,S.T., Chu,M., Kim,H., Joo,S., Kim,C., Song,W. and Yoo,H.  
BAC end sequences of Library RP-43  
Unpublished  
2 (bases 1 to 362)  
Park,H., Kim,Y., Kim,S., Han,Y., Woo,T., Park,K., Eun,C.J.,  
Hoon,S.T., Chu,M., Kim,H., Joo,S., Kim,C., Song,W. and Yoo,H.  
Submitted (07-JAN-2002) Hong-Seog Park, Korea Research Institute of  
Direct Submission  
Bioscience and Biotechnology (KRIIBB), Genome Research Center (GRC);  
52, Oun-dong, Yuseong-gu, Daejeon 305-333, Korea  
(E-mail:redstoneemail.kribb.re.kr, URL:http://phs.grc.kribb.re.kr/,  
Tel:82-42-866-7181, Fax:82-42-860-4409)  
Clones are derived from the chimpanzee BAC library RP-43 This BAC  
end was generated during the R&D process and may have higher chance  
of clone tracking errors.  
PRIMERS  
Sequencing: TJ  
LIBRARY  
Vector : pBACe3.6  
R.Site 1 : EcoRI  
R.Site 2 : EcoRI  
Location/Qualifiers  
1 .362  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
/clone="RP43-088K24.TJ"  
/sex="male"  
/cell\_type="lymphocytes"  
/clone\_lib="RP-43 Chimpanzee Male BAC Library"

```

ORIGIN
Query Match          92.0%; Score 18.4; DB 10; Length 362;
Best Local Similarity 95.0%; Pred. No. 6.9e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGACCAAC 20
    |||||
Db 231 GAGCCAAACAGGACCAAC 212

RESULT 2
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Oryza coarctata (Porteresia coarctata)
Oryza coarctata
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE
AUTHORS
Kim.H., Collura,K., Wisotski,M., Byrne,M., Stum,D., Smart,D.,
Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C., Soderlund,C. and
Wing,R.
TITLE
JOURNAL
COMMENT
OMAP (Oryza Map Alignment Project) - Arizona Genomics Institute
Unpublished (2005)
Contact: Rod A. Wing
Arizona Genomics Institute
University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: twing@genome.arizona.edu
PCR Primers
FORWARD: TAA TAC GAC TCA CTA TAG GG
BACKWARD: CAC TCA TTA GGC ACC CCA
Plate: 0180 row: F column: 03
Seq primer: TAA TAC GAC TCA CTA TAG GG
Class: BAC ends.
FEATURES
source
1..752
/organism="Oryza coarctata"
/mol_type="genomic DNA"
/db_xref="taxon:77588"
/clone="OC_Ba0180F03"
/tissue_type="leaves"
/dev_stage="mature"
/lab_host="DH10B"
/clone_lib="OC_Ba"
/note="Vector: pAG1BAC1; Site_1: HindIII; Site_2: HindIII"

ORIGIN
Query Match          92.0%; Score 18.4; DB 10; Length 752;
Best Local Similarity 95.0%; Pred. No. 7.8e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGACCAAC 20
    |||||
Db 91 GAGCCAAACAGGACCAAC 110

RESULT 3
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
EST.

SOURCE
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
REFERENCE
AUTHORS
NIH-MGC http://mgc.nci.nih.gov/.
TITLE
JOURNAL
COMMENT
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9696 row: c column: 03
High quality sequence stop: 559.
Location/Qualifiers
1..938
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3899066"
/tissue_type="epithelioid carcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC_70"
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.1 kb. Library constructed by Life
Technologies."

ORIGIN
Query Match          92.0%; Score 18.4; DB 2; Length 938;
Best Local Similarity 95.0%; Pred. No. 8.1e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGACCAAC 20
    |||||
Db 852 GAGCCAAACAGGACCAAC 871

RESULT 4
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Muridae; Murinae; Mus.
REFERENCE
AUTHORS
Aizawa,K., Akahira,S., Akimura,T., Arai,A., Arakawa,T.,
Carninci,P., Hanagaki,T., Hayatsu,N., Hiraoka,T., Hirozane,T.,
Hodoyama,Y., Imotani,K., Iehi,Y., Itoh,M., Izawa,M., Kawai,J.,
Kojima,Y., Konno,H., Kusakabe,M., Matsuura,T., Miyazaki,A.,
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Okazaki,Y.,
Okido,T., Owa,C., Sakai,K., Sakai,K., Sasaki,D., Sato,K.,
Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y.,
Suzuki,H., Tagawa,A., Takahashi,F., Tanaka,T., Toya,T.,
Watahiki,A., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshiki,A.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Aizawa,K. et al. 2000)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute

```

The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216

Email: genome-res@sc.riken.jp, URL: <http://genome.gsc.riken.jp/>  
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S.,  
Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Thermotabilization and thermoactivation of thermolabile enzymes by  
trehalose and its application for the synthesis of full length  
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Itoh, M., Kusunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,  
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M.,  
Okazaki, Y. and Hayashizaki, Y.  
Automated filtration-based high-throughput plasmid preparation  
system. Genome Res. 9 (5), 463-470 (1999)  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,  
19-44 (1999)  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for  
further details.

#### FEATURES

Location/Qualifiers  
1. .300  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="2200001A16"  
/sex="male"  
/tissue\_type="stomach"  
/dev\_stage="adult"  
/lab\_host="SOLR"  
/clone\_lib="RIKEN full-length enriched, adult male  
stomach"  
/notes="Site 1: XhoI; Site 2: SstI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5'  
GAGAGAGAGCGGCGCAACTCGAGTTTCTTTTCTTTT 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. Second strand cDNA was prepared with the  
primer adapter of sequence [5'  
GAGAGAGATTCTCGAGTTAATTAATTAATTCCTCCCTCCCTCCCTCC 3']. cDNA  
was cleaved with XhoI and SstI."

#### ORIGIN

Query Match 87.0%; Score 17.4; DB 2; Length 300;  
Best Local Similarity 94.7%; Pred. No. 2e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GAGCCAAACAGGACCAAA 19  
|||||  
DB 149 GAGCCAAAGGAACCAAA 167

RESULT 5  
BQ329066  
LOCUS BQ329066 337 bp mRNA linear EST 17-MAY-2002  
DEFINITION CM3-EN0047-031100-460-a09 EN0047 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BQ329066  
VERSION BQ329066.1 GI:20969175  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
1 (bases 1 to 337)

REFERENCE  
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,

#### TITLE

JOURNAL  
PUBMED  
COMMENT

Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V.,  
O'Hare, M.J., Soares, P., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
10737800  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM3&t2=CM3-EN0047-031100-460-a09&t3=2000-11-03&t4=1>)  
Seq primer: puc 18 forward  
High quality sequence start: 18  
High quality sequence stop: 333.

#### FEATURES

Location/Qualifiers  
1. .337  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="EN0047"  
/note="Organ: lung normal; Vector: puc18; Site 1: SmaI;  
Site 2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

#### ORIGIN

Query Match 87.0%; Score 17.4; DB 3; Length 337;  
Best Local Similarity 94.7%; Pred. No. 2e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 AGCCAAACAGGACCAAAAC 20  
|||||  
DB 313 AGCCCAACAGGACCAAAAC 331

#### RESULT 6

BH288892  
LOCUS BH288892 357 bp DNA linear GSS 30-NOV-2001  
DEFINITION CH230-2L10-TV CHORI-230 Segment 1 Rattus norvegicus genomic clone  
CH230-2L10, genomic survey sequence.  
ACCESSION BH288892  
VERSION BH288892.1 GI:17201300  
KEYWORDS GSS.  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridea; Muridae; Murinae; Rattus.  
1 (bases 1 to 357)

#### REFERENCE

AUTHORS Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K.,  
Shvartabeyn, A., Gebregorgis, E., Overton, L., Russell, D., Chen, D.,  
Riggs, F., de Jong, P. and Fraser, C.M.  
Rat BAC End Sequences from Library CHORI-230 EcoRI segment  
Unpublished (1999)

#### TITLE

JOURNAL  
COMMENT  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208

Email: szhao@tigr.org  
 Clones are derived from the rat BAC library CHORI-230  
 (<http://www.chori.org/bacpac/rat230.htm>). For BAC library  
 availability, please contact Pieter de Jong ([pdejong@mail.cho.org](mailto:pdejong@mail.cho.org)).  
 Clones may be purchased from BACPAC Resources  
 (<http://www.chori.org/bacpac/oreringinformation.html>). BAC end  
 page: [http://www.tigr.org/tdb/bac\\_ends/rat/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html)  
 Plate: 2 row: L column: 10  
 Seq primer: T7  
 Class: BAC ends.

#### FEATURES

source  
 Location/Qualifiers  
 1. .357  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /strain="BN/SsNHsd/MCW"  
 /db\_xref="taxon:10116"  
 /clone="CH230-2L10"  
 /sex="Female"  
 /cell\_type="Brain"  
 /clone\_lib="CHORI-230 Segment 1"  
 /note="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;  
 CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by  
 Pieter de Jong"

#### ORIGIN

Query Match 87.0%; Score 17.4; DB 9; Length 357;  
 Best Local Similarity 94.7%; Pred. No. 2e+03;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GAGCCAAACAGGAGCCAAA 19  
 |||||  
 Db 227 GAGCCAAACAGGAGCCAAA 245

#### RESULT 7

AW080511/c  
 LOCUS  
 DEFINITION  
 xc24a11.x1 NCI CGAP Col9 Homo sapiens cDNA clone IMAGE:2585180 3,  
 similar to confains Alu repetitive element; mRNA sequence.  
 AW080511  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homnidae; Homo.  
 1 (bases 1 to 386)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)  
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael  
 R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: Life  
 Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The  
 I.M.A.G.E. Consortium DNA Sequencing by: Washington University  
 Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

#### FEATURES

source  
 Location/Qualifiers  
 1. .386  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2585180"  
 /tissue\_type="moderately differentiated adenocarcinoma"  
 Possible reversed clone: polyT not found  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 378.

/lab\_host="DH10B"  
 /clone\_lib="NCI CGAP Col9"  
 /note="Organ: Colon; Vector: pCMV-SPORT6; Site 1: SalI;  
 Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
 Normalized to Cot 50. Average insert size 1.32kb.  
 Normalized version of NCI CGAP Col9. Library constructed  
 by Life Technologies."

#### ORIGIN

Query Match 87.0%; Score 17.4; DB 1; Length 386;  
 Best Local Similarity 94.7%; Pred. No. 2.1e+03;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GAGCCAAACAGGAGCCAAA 19  
 |||||  
 Db 321 GAGCCAAACAGGAGCCAAA 303

#### RESULT 8

CG363310/c  
 LOCUS  
 DEFINITION  
 CG363310 413 bp DNA linear GSS 26-AUG-2003  
 genomic survey sequence.  
 CG363310  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Zea mays  
 Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Zea.  
 1 (bases 1 to 413)  
 Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,  
 Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,  
 Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.  
 Consortium for Maize Genomics  
 Unpublished (2002)  
 Other\_GSSs: CGWGD52TV  
 Contact: Cathy Whitelaw  
 TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-838-5843  
 Fax: 301-838-0208  
 Email: [whitelaw@tigr.org](mailto:whitelaw@tigr.org)  
 Seq primer: TR  
 Class: methylation filtered.  
 Location/Qualifiers  
 1. .413  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"  
 /clone="ZMMBMA0606708"  
 /clone\_lib="ZM 0.7 1.5 KB"  
 /note="Vector: pBGSK-; Site 1: HincII; 0.7-1.5 kb  
 methylation filtered genomic DNA library"

#### FEATURES

source  
 Location/Qualifiers  
 1. .413  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"  
 /clone="ZMMBMA0606708"  
 /clone\_lib="ZM 0.7 1.5 KB"  
 /note="Vector: pBGSK-; Site 1: HincII; 0.7-1.5 kb  
 methylation filtered genomic DNA library"

#### ORIGIN

Query Match 87.0%; Score 17.4; DB 10; Length 413;  
 Best Local Similarity 94.7%; Pred. No. 2.1e+03;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 AGCCAAACAGGAGCCAAAC 20  
 |||||  
 Db 364 AGCCAAACAGGAGCCAAAC 346

#### FEATURES

source  
 Location/Qualifiers  
 9  
 BH262971  
 LOCUS  
 DEFINITION  
 CH230-121C5.TUB CHORI-230 Segment 1 Rattus norvegicus genomic clone  
 CH230-121C5, genomic survey sequence.  
 ACCESSION  
 BH262971

```

VERSION BH262971.1 GI:17170170
KEYWORDS GSS.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 414)
AUTHORS Zhao,S., Shetty,J., Shatsman,S., Teegave,G., Geer,K.,
Shvartsbeyn,A., Gebregorgis,E., Overton,L., Russell,D., Chen,D.,
Riggs,F., de Jong,P. and Frazer,C.M.
Rat BAC End sequences from Library CHORI-230 EcoRI segment
Unpublished (1999)
Other GSSs: CH230-121C5.TVB
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat230.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or ering.information.htm). BAC end
page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Plate: 121 row: C column: 5
Seq primer: SP6
Class: BAC ends.
FEATURES
    Location/Qualifiers
        1..414
            /organism="Rattus norvegicus"
            /mol_type="genomic DNA"
            /strain="BN/SENHsd/MCW"
            /db_xref="taxon:10116"
            /clone="CH230-121C5"
            /sex="Female"
            /cell_type="Brain"
            /notes="Vector: PTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;
            CHORI-230 Rat (BN/SENHsd/MCW) BAC library produced by
            Pieter de Jong"
ORIGIN
    Query Match 87.0% Score 17.4; DB 9; Length 414;
    Best Local Similarity 94.7%; Pred. No. 2.1e+03;
    Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAA 19
    |||||
Db 263 GAGCCAAACAGGAGCCAA 281

RESULT 10
BY692431/c
LOCUS BY692431 RIKEN full-length enriched, osteoclast-like cell Mus
DEFINITION musculus cDNA clone I420022016 3', mRNA sequence.
ACCESSION BY692431
VERSION BY692431.1 GI:27101887
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 418)
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,

```

```

Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
Gustinich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomika,M.,
Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
Rogers,J., Birney,E. and Hayashizaki,Y.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
12466851
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gs.riken.jp, URL:http://genome.gsc.riken.jp/
Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,
Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,
Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,
Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,
Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and
Hayashizaki,Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Tissues were provided by Takashi Ishikawa ( Department of Surgery
2 Yokohama City University 3-9 Fukuura,Kanazawa-ku,Yokohama
236-0004 Japan ) whose assistance we gratefully acknowledge.
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
FEATURES
    Location/Qualifiers
        1..418
            /organism="Mus musculus"
            /mol_type="mRNA"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="I420022016"
            /cell_type="osteoclast-like cell"
            /clone_lib="RIKEN full-length enriched, osteoclast-like
            cell"
ORIGIN

```

```

Query Match      87.0%; Score 17.4; DB 5; Length 418;
Best Local Similarity 94.7%; Pred. No. 2.1e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAA 19
Db 354 GAGCAAAACAGGAACCAA 336

RESULT 11
AI672491
LOCUS
DEFINITION
  wa03h01.x1 NCI CGAP Kid11 Homo sapiens cDNA clone IMAGE:2297041 3',
  similar to contains element MER30 repetitive element ;, mRNA
  sequence.
ACCESSION
  AI672491
VERSION
  AI672491.1 GI:4852222
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
  Homnidae; Homo.
REFERENCE
  1 (bases 1 to 430)
  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
  Tumor Gene Index
  Unpublished (1997)
JOURNAL
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs-r@mail.nih.gov
  Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
  Emmert-Buck, M.D., Ph.D.
  cDNA Library Preparation: M. Bento Soares, Ph.D.
  cDNA Library Arrayed by: Greg Lennon, Ph.D.
  DNA Sequencing by: Washington University Genome Sequencing Center
  Cloned through the NCI-CGAP clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at:
  www.bio.llnl.gov/bbrp/image/image.html
  Insert Length: 972 Std Error: 0.00
  Seq primer: -40UP from Gibco
  High quality sequence stop: 426.
  Location/Qualifiers
  1..430
  /organism="Homo sapiens"
  /mol_type="mRNA"
  /db_xref="taxon:9606"
  /clone="IMAGE:2297041"
  /lab_host="DH10B"
  /clone_lib="NCI CGAP Kid11"
  /note="Organ: Kidney; Vector: pT73D-Pac (Pharmacia) with
  a modified polylinker; Site: 1: Not I; Site 2: Eco RI;
  Plasmid DNA from the normalized library NCI CGAP Kid3 was
  prepared, and ss circles were made in vitro. Following HAP
  purification, this DNA was used as tracer in a subtractive
  hybridization reaction. The driver was PCR-amplified cDNAs
  from a pool of 5,000 clones made from the same library
  (clonids 1322376-1323911, 1456007-1456775, and
  1500552-1502855). Subtraction by Bento Soares and M.
  Fatima Bonaldo. "

ORIGIN
Query Match      87.0%; Score 17.4; DB 1; Length 430;
Best Local Similarity 94.7%; Pred. No. 2.1e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAA 19
Db 109 GAGCCCAACAGGAACCAA 127

RESULT 12
AA607599
LOCUS
DEFINITION
  Barstead mouse irradiated colon MPLRB7 Mus musculus cDNA
  clone IMAGE:1053152 5', mRNA sequence.
ACCESSION
  AA607599
VERSION
  AA607599.1 GI:2455034
KEYWORDS
  EST.
SOURCE
  Mus musculus (house mouse)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
  Sciurognathi; Muridea; Muridae; Murinae; Mus.
REFERENCE
  1 (bases 1 to 463)
  Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
  Geisels, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
  Schellenberg, K., Steptoe, M., Tan, P., Underwood, K., Moore, B.,
  Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
  Waterston, R.
  The WashU-HHMI Mouse EST Project
  Unpublished (1996)
JOURNAL
  Contact: Marra M/Mouse EST Project
  WashU-HHMI Mouse EST Project
  Washington University School of Medicine
  4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
  Tel: 314 286 1800
  Fax: 314 286 1810
  Email: mouseest@watson.wustl.edu
  This clone is available royalty-free through LLNL; contact the
  IMAGE Consortium (info@image.llnl.gov) for further information.
  MGI:584728
  Seq primer: -28mi3 rev2 ET from Amersham
  High quality sequence stop: 439.
  Location/Qualifiers
  1..463
  /organism="Mus musculus"
  /mol_type="mRNA"
  /strain="FVB/N"
  /db_xref="taxon:10090"
  /clone="IMAGE:1053152"
  /dev_stage="8 weeks"
  /lab_host="DH10B"
  /clone_lib="Barstead mouse irradiated colon MPLRB7"
  /note="Vector: pT73D-Pac (Pharmacia) with a modified
  polylinker; Site 1: EcoRI; Site 2: NotI; Tissue obtained
  from 8 week old mouse. Colon was harvested 72 hours after
  irradiation with 1400 Gys. 1st strand cDNA was primed
  with a Not I - oligo(dT) primer
  15'TGTACGATCTGAGTGGAGCGCGCCCTTTTCTTTTCTTTTCTTTTCTTTT
  T 3'; double-stranded cDNA was ligated to Eco RI
  adaptors [AATTCGATCCTTG], digested with Not I and cloned
  into the Not I and Eco RI sites of the modified pT7T3
  vector. Library constructed by Bob Barstead. "

ORIGIN
Query Match      87.0%; Score 17.4; DB 1; Length 463;
Best Local Similarity 94.7%; Pred. No. 2.1e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAA 19
Db 34 GAGCCAAACAGGAACCAA 52

RESULT 13
CB161822/c
LOCUS
DEFINITION
  K-EST0221836 L17N670205nl Homo sapiens cDNA clone
  L17N670205nl-50-D12 5', mRNA sequence.
ACCESSION
  CB161822
VERSION
  CB161822.1 GI:28147948
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

```



REFERENCE  
AUTHORS  
Hominidae; Homo.  
1. (Bases 1 to 477)  
Kim, N. S., Hahn, Y., Oh, J. H., Lee, J. Y., Ahn, H. Y., Chu, M. Y., Kim, M. R., Oh, K. J., Cheong, J. E., Sohn, H. Y., Kim, J. M., Park, H. S., Kim, S. and Kim, Y. S.

## ORIGIN

```
Query Match      87.0%; Score 17.4; DB 6; Length 477;
Best Local Similarity 94.7%; Pred. NO. 2.1e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```

/mol_type="mENA"
/db_xref="taxon.9913"
/clone="KN227-028.117"
/tissue_type="uterus"
/clone_lib="KN224"
/notes="vector: pBluescriptII(SK+); Site 1: EcoRI; Site 2:
NotI; single pass sequencing. Normalised library
constructed from Bovine Uterus tissue."

```

## ORIGIN

Query Match 87.0%; Score 17.4; DB 1; Length 491;  
Best Local Similarity 94.7%; Pred. No. 2.2e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

## ORIGIN

Query Match 87.0%; Score 17.4; DB 2; Length 492;

```

Best Local Similarity 94.7%; Pred. No. 2.2e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGCCAAACAGGAACCAAC 20
    |||||
Db 293 AGCCAAACAGGAACCAAC 275
    |||||

RESULT 16
BZ634760/c
LOCUS BZ634760 545 bp DNA linear GSS 29-JAN-2003
DEFINITION PUA668TD ZM 0.6 1.0 KB Zea mays genomic clone ZMBTa001K15,
genomic survey sequence.
ACCESSION BZ634760
VERSION BZ634760.1 GI:28082274
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
Clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 545)
REFERENCE
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
Bennetzen,J.
Maize Genomics Consortium
Unpublished (2003)
Other_GSSs: PUA668TB
Contact: Cathy Whitelaw
TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.

FEATURES
source
1..545
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMBTa001K15"
/clone_lib="ZM 0.6 1.0 KB"
/notes="Vector: PCR4-TOFO; Site_1: EcoRI; 0.6-1.0 kb high
COT selected genomic DNA library"

ORIGIN
Query Match 87.0%; Score 17.4; DB 9; Length 545;
Best Local Similarity 94.7%; Pred. No. 2.2e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGCCAAACAGGAACCAAC 20
    |||||
Db 88 AGCCAAACAGGAATCAAC 70
    |||||

RESULT 17
CB157333/c
LOCUS CB157333 549 bp mRNA linear EST 29-JAN-2003
DEFINITION K-EST0216437 L17N670205n1 Homo sapiens cDNA clone
L17N670205n1-3-C01 5', mRNA sequence.
ACCESSION CB157333
VERSION CB157333.1 GI:28142465
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 549)
REFERENCE
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.

FEATURES
source
1..549
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="L17N670205n1-43-E01"
/sex="F"

```

```

Kim,Y.S.
21C Frontier Korean EST Project 2001
Unpublished (2002)
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 3 row: C column: 01
High quality sequence stop: 549.

FEATURES
source
1..549
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="L17N670205n1-3-C01"
/sex="F"
/lab_host="Top10P"
/clone_lib="L17N670205n1"
/notes="Organ: Liver; Vector: p7T3-Pac; Site_1: EcoRI;
Site_2: NotI; The library was contributed by the Soares
Laboratory and it was constructed as described by Bonaldo,
M.F., Lennon, G. and Soares, M.B. (1996), Genome Research
6(9): 791-806. RNA was prepared from harvested cell
culture."

ORIGIN
Query Match 87.0%; Score 17.4; DB 6; Length 549;
Best Local Similarity 94.7%; Pred. No. 2.2e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGCCAAACAGGAACCAAC 20
    |||||
Db 61 AGCCCAACAGGAACCAAC 43
    |||||

RESULT 18
CB164700/c
LOCUS CB164700 549 bp mRNA linear EST 30-JAN-2003
DEFINITION K-EST0225965 L17N670205n1 Homo sapiens cDNA clone
L17N670205n1-43-E01 5', mRNA sequence.
ACCESSION CB164700
VERSION CB164700.1 GI:28150826
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 549)
REFERENCE
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
21C Frontier Korean EST Project 2001
Unpublished (2002)
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 43 row: E column: 01
High quality sequence stop: 549.

FEATURES
source
1..549
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="L17N670205n1-43-E01"
/sex="F"

```



```

source
1. .554
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMBT001K15"
/clone_lib="ZM_0.6_1.0 KB"
/notes="Vector: PCR4-TOFO; Site_1: EcoRI; 0.6-1.0 kb high
cor selected genomic DNA library"

ORIGIN
Query Match      87.0%; Score 17.4; DB 10; Length 554;
Best Local Similarity 94.7%; Pred. No. 2.2e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 AGCCAAACAGGAACCAAC 20
Db      458 AGCCAAACAGGAATCAAC 476

RESULT 22
CW378022
LOCUS
DEFINITION
Sorghum bicolor genomic clone f585b001f056j16, genomic survey
sequence.
ACCESSION
CW378022
VERSION
CW378022.1 GI:55096466
KEYWORDS
Sorghum bicolor (sorghum)
SOURCE
Sorghum bicolor
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE
1 (bases 1 to 585)
Bedell J.A., Budiman, M.A., Nunberg, A., Citek, R.W., Robbins, D.,
Jones, J., Flick, E., Rohlfing, T., Fries, J., Bradford, K.,
McMenamy, J., Smith, M., Holeman, H., Roe, B.A., Wiley, G., Korff, I.F.,
Rabinowicz, P.D., Lakey, N., McCombie, W.R., Jeddleloh, J.A. and
Martienssen, R.A.
Sorghum genome sequencing by methylation filtration
PLOS Biol. 3 (1), e13 (2005)
15660154
Contact: Bedell JA
Orion Genomics, LLC
4041 Forest Park Ave, St. Louis, MO 63108, USA
Tel: 314 615 6979
Fax: 314 615 5975
Email: jbedell@oriongenomics.com
Plate: f585b001f056 row: j column: 16
Seq primer: f Forward
Class: methylation filtered
High quality sequence stop: 585.
Location/Qualifiers
1. .595
/organism="Sorghum bicolor"
/mol_type="genomic DNA"
/cultivar="Atx623"
/db_xref="taxon:4558"
/clone="f585b001f056j16"
/clone_lib="Sorghum methylation filtered library (LibID:
104)"
/notes="Organ: leaf; Vector: pBCSK(-); Site_1: HincII; DNA
prepared from purified nuclei was randomly sheared,
end-repaired, size fractionated to enrich for the 0.5 to 5
kb fraction, ligated into HincII-digested pBCSK(-) vector
and electroporated into E. coli cells. This is a
methylation filtered library."

ORIGIN
Query Match      87.0%; Score 17.4; DB 10; Length 585;
Best Local Similarity 94.7%; Pred. No. 2.2e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 GAGCCAAACAGGAACCAAA 19
Db      118 GAGCCAAACATCAACCAAA 136

RESULT 23
AU019801/c
LOCUS
DEFINITION
AU019801 Mouse eight-cell stage embryo cDNA Mus musculus cDNA clone
J0522F05 3', mRNA sequence.
ACCESSION
AU019801
VERSION
AU019801.1 GI:3375385
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 588)
Ko, M.S.H., Kitchen, J.R., Wang, X., Threat, T.A., Sun, T., DePalma, G.E.,
Liang, Y., Kargul, G.J., Sharara, R. and Doi, H.
Systematic analyses of genes expressed in eight-cell stage mouse
embryos (The ERATO/Doi Project at Wayne State University)
(Ko, M.S.H. et al.)
Unpublished (1998)
JOURNAL
Contact: Hirofumi Doi
COMMENT
Doi Bioasymmetry Project, ERATO
Japan Science and Technology Corporation (JST)
WBG Marive East 12F, 2-6 Nakase, Mihama-ku, Chiba 261-71, Japan
Email: hdoibioa.jst.go.jp.
Location/Qualifiers
1. .588
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="J0522F05"
/dev_stage="eight-cell stage embryo"
/clone_lib="Mouse eight-cell stage embryo cDNA"

ORIGIN
Query Match      87.0%; Score 17.4; DB 1; Length 588;
Best Local Similarity 94.7%; Pred. No. 2.2e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 AGCCAAACAGGAACCAAAAC 20
Db      490 AGCCAAACAGGAACCGAAC 472

RESULT 24
BB631472
LOCUS
DEFINITION
BB631472 RIKEN full-length enriched, 16 days neonate thymus Mus
musculus cDNA clone A130079F14 5', mRNA sequence.
ACCESSION
BB631472
VERSION
BB631472.1 GI:16468219
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 628)
Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T.,
Hara, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J.,
Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,
Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Arakawa, T., et al. 2001)

```



**AUTHORS**  
Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K., Shvartsbeyn, A., Gebregeorgis, E., Overton, L., Russell, D., Chen, D., Riggs, F., de Jong, P. and Fraser, C.M.  
**TITLE**  
Rat BAC End Sequences from Library CHORI-230 MboI segment  
**JOURNAL**  
**COMMENT**  
Other GSSs: CH230-361C5.TV  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the rat BAC library CHORI-230  
(<http://www.chori.org/bacpac/rat230.htm>). For BAC library availability, please contact Pieter de Jong ([pdejong@mail.cho.org](mailto:pdejong@mail.cho.org)). Clones may be purchased from BACPAC Resources ([http://www.chori.org/bacpac/or\\_ering\\_information.htm](http://www.chori.org/bacpac/or_ering_information.htm)). BAC end page: [http://www.tigr.org/tadb/bac\\_ends/rat/bac\\_end\\_intro.html](http://www.tigr.org/tadb/bac_ends/rat/bac_end_intro.html)  
Plate: 361 row: C column: 5  
Seq primer: SP6  
Class: BAC ends.  
**FEATURES**  
source  
Location/Qualifiers  
1. .660  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/strain="BN/SsNHsd/MCW"  
/db\_xref="taxon:10116"  
/clone="CH230-361C5"  
/sex="female"  
/cell\_type="Brain"  
/clone\_lib="CHORI-230 Segment 2"  
/note="Vector: pTARBAC1.3; Site 1: MboI; Site 2: MboI; CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by Pieter de Jong"

**ORIGIN**  
Query Match 87.0%; Score 17.4; DB 9; Length 660;  
Best Local Similarity 94.7%; Pred. No. 2.3e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

**QY** 1 GAGCCAAACAGGAGCCAAA 19  
|||||  
**Db** 283 GAGCCAAACAGGAGCCAAA 301  
|||||

**RESULT 27**  
CG852549/c  
**LOCUS** ZMMBB0346J19.f ZMMBB Zea mays genomic clone ZMMBB0346J19 5', genomic survey sequence.  
**DEFINITION**  
**ACCESSION** CG852549  
**VERSION** CG852549.1 GI:38379410  
**KEYWORDS** GSS.  
**SOURCE** Zea mays  
**ORGANISM**  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.  
**REFERENCE** 1 (bases 1 to 666)  
**AUTHORS** Yu, Y., Kim, H.R., Hatfield, J., Soderlund, C., Bharti, A.K., Messing, J. and Wing, R.  
**TITLE** Sequencing of the maize genome  
**JOURNAL** Unpublished (2003)  
**COMMENT** Contact: Rod Wing  
Arizona Genomics Institute  
University of Arizona  
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ 85721-0088, USA  
Tel: 520 626 3967  
Fax: 520 621 9288  
Email: [rwing@genome.arizona.edu](mailto:rwing@genome.arizona.edu)  
PCR Primers

**AUTHORS**  
Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K., Shvartsbeyn, A., Gebregeorgis, E., Overton, L., Russell, D., Chen, D., Riggs, F., de Jong, P. and Fraser, C.M.  
**TITLE**  
Rat BAC End Sequences from Library CHORI-230 MboI segment  
**JOURNAL**  
**COMMENT**  
Other GSSs: CH230-361C5.TV  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the rat BAC library CHORI-230  
(<http://www.chori.org/bacpac/rat230.htm>). For BAC library availability, please contact Pieter de Jong ([pdejong@mail.cho.org](mailto:pdejong@mail.cho.org)). Clones may be purchased from BACPAC Resources ([http://www.chori.org/bacpac/or\\_ering\\_information.htm](http://www.chori.org/bacpac/or_ering_information.htm)). BAC end page: [http://www.tigr.org/tadb/bac\\_ends/rat/bac\\_end\\_intro.html](http://www.tigr.org/tadb/bac_ends/rat/bac_end_intro.html)  
Plate: 361 row: C column: 5  
Seq primer: SP6  
Class: BAC ends.  
**FEATURES**  
source  
Location/Qualifiers  
1. .660  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/strain="BN/SsNHsd/MCW"  
/db\_xref="taxon:10116"  
/clone="CH230-361C5"  
/sex="female"  
/cell\_type="Brain"  
/clone\_lib="CHORI-230 Segment 2"  
/note="Vector: pTARBAC1.3; Site 1: MboI; Site 2: MboI; CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by Pieter de Jong"

**ORIGIN**  
Query Match 87.0%; Score 17.4; DB 9; Length 660;  
Best Local Similarity 94.7%; Pred. No. 2.3e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

**QY** 1 GAGCCAAACAGGAGCCAAA 19  
|||||  
**Db** 283 GAGCCAAACAGGAGCCAAA 301  
|||||

**RESULT 27**  
CG852549/c  
**LOCUS** ZMMBB0346J19.f ZMMBB Zea mays genomic clone ZMMBB0346J19 5', genomic survey sequence.  
**DEFINITION**  
**ACCESSION** CG852549  
**VERSION** CG852549.1 GI:38379410  
**KEYWORDS** GSS.  
**SOURCE** Zea mays  
**ORGANISM**  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.  
**REFERENCE** 1 (bases 1 to 666)  
**AUTHORS** Yu, Y., Kim, H.R., Hatfield, J., Soderlund, C., Bharti, A.K., Messing, J. and Wing, R.  
**TITLE** Sequencing of the maize genome  
**JOURNAL** Unpublished (2003)  
**COMMENT** Contact: Rod Wing  
Arizona Genomics Institute  
University of Arizona  
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ 85721-0088, USA  
Tel: 520 626 3967  
Fax: 520 621 9288  
Email: [rwing@genome.arizona.edu](mailto:rwing@genome.arizona.edu)  
PCR Primers

**FORWARD:** T7  
**BACKWARD:** M13  
Plate: 0346 row: J column: 19  
Seq primer: T7  
Class: BAC ends.  
**FEATURES**  
Location/Qualifiers  
1. .666  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/cultivar="B73"  
/db\_xref="taxon:4577"  
/clone="ZMMBB0346J19"  
/lab\_host="DH10B"  
/clone\_lib="ZMMBB"  
/note="Vector: pBeloBAC11; Site 1: HindIII; Site 2: HindIII; Zea mays L. sep. mays"  
**ORIGIN**  
Query Match 87.0%; Score 17.4; DB 10; Length 666;  
Best Local Similarity 94.7%; Pred. No. 2.3e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

**QY** 2 AGCCAAACAGGAACCAAC 20  
|||||  
**Db** 423 AGCCAAACAGGAATCAAC 405  
|||||

**RESULT 28**  
CR887634  
**LOCUS** Sus scrofa BES, genomic survey sequence.  
**DEFINITION** CR887634  
**ACCESSION** CR887634.1 GI:55966298  
**VERSION** GSS; BAC-end sequence BES; Genome Survey Sequence.  
**KEYWORDS** Sus scrofa (pig)  
**SOURCE** Sus scrofa  
**ORGANISM**  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae; Sus.  
**REFERENCE** 1 (bases 1 to 666)  
**AUTHORS** Rogel-Gaillard, C., Bourgeaux, N., Billault, A., Vaiman, M. and Chardon, P.  
**TITLE** Construction of a swine BAC library: application to the characterization and mapping of porcine type C endoviral elements  
**JOURNAL** Cytogenet. Cell Genet. 85 (3-4), 205-211 (1999)  
**PUBMED** 10449899  
**REFERENCE** 2 (bases 1 to 666)  
**AUTHORS** Chardon, P., Iannuccelli, N., Roig, A., Dossat, C., Demars, J., Rogel-Gaillard, C., Roy, A., Schibler, L. and Milan, D.  
**TITLE** A physical map of the swine genome  
**JOURNAL** Unpublished  
**REFERENCE** 3 (bases 1 to 666)  
**AUTHORS** Genoscope.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (18-NOV-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : [seqref@genoscope.cns.fr](mailto:seqref@genoscope.cns.fr))  
- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
**FEATURES**  
Location/Qualifiers  
1. .666  
/organism="Sus scrofa"  
/mol\_type="genomic DNA"  
/strain="Large White"  
/db\_xref="taxon:9823"  
/clone="b10125A08"  
/sex="male"  
/cell\_type="fibroblast"  
/clone\_lib="SBAB"  
/note="Genoscope sequence ID : IH0AAA10DA05FMI"  
**ORIGIN**  
Query Match 87.0%; Score 17.4; DB 11; Length 666;  
Best Local Similarity 94.7%; Pred. No. 2.3e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;



9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-838-5843  
 Fax: 301-838-0208  
 Email: whitelaw@tigr.org  
 Seq primer: TF  
 Class: sheared ends.

## FEATURES

## source

1. .777  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"  
 /clone="ZMMBTa175E22"  
 /clone\_lib="ZM\_0.6-1.0 KB"  
 /note="Vector: PCR4-TOPO; Site\_1: EcoRI; 0.6-1.0 kb high  
 Cot selected genomic DNA library"

## ORIGIN

Query Match 87.0%; Score 17.4; DB 9; Length 777;  
 Best Local Similarity 94.7%; Pred. No. 2.3e+03;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGCCAAACAGGACCAAC 20  
 |||||  
 DB 503 AGCCAAACAGGACCAAGC 521

## RESULT 32

## CC750285/c

## LOCUS

DEFINITION ZMMBBb0130L20\_r ZMMBBb Zea mays genomic clone ZMMBBb0130L20 3',  
 genomic survey sequence.

## ACCESSION

## CC750285

## VERSION

CC750285.1 GI:32208194

## KEYWORDS

## GSS.

## SOURCE

## Zea mays

## Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Zea.  
 1 (bases 1 to 781)

## REFERENCE

## AUTHORS

Yu, Y., Kim, H.R., Hatfield, J., Soderlund, C., Bharti, A.K., Messing, J.  
 and Wing, R.

## TITLE

## Sequencing of the maize genome

## Unpublished (2003)

## CONTACT: Rod Wing

## Arizona Genomics Institute

## University of Arizona

## Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ

## 85721-0088, USA

## Tel: 520 626 3967

## Fax: 520 621 9288

## Email: rwing@genome.arizona.edu

## PCR Primers

## FORWARD: T7

## BACKWARD: M13r

## Plate: 0130 row: L column: 20

## Seq primer: M13r

## Class: BAC ends.

## FEATURES

## source

1. .781  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /cultivar="B73"  
 /db\_xref="taxon:4577"  
 /clone="ZMMBBb0130L20"  
 /lab\_host="DH10B"  
 /clone\_lib="ZMMBBb"  
 /note="Vector: pBelobAC11; Site\_1: HindIII; Site\_2:  
 HindIII; Zea mays L. ssp. mays"

## ORIGIN

Query Match 87.0%; Score 17.4; DB 9; Length 781;

Best Local Similarity 94.7%; Pred. No. 2.3e+03;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGCCAAACAGGACCAAC 20  
 |||||  
 DB 88 AGCCAAACAGGACCAAC 70

## RESULT 33

## DR474762

## LOCUS

## DEFINITION

WS00962.C21 M09 IS-B-N-A-10 Picea engelmannii x Picea sitchensis  
 cDNA clone WS00962\_M09 3, mRNA sequence.

## ACCESSION

## DR474762

## VERSION

## EST.

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

Picea engelmannii x Picea sitchensis  
 Picea engelmannii x Picea sitchensis  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.  
 1 (bases 1 to 791)  
 Ralph, S., Kolosova, N., Oddy, C., Cooper, D., Butterfield, Y.,  
 Kirkpatrick, R., Liu, J., Palmquist, D., Stott, J., Barber, S., Yang, G.,  
 Babakaiff, R., Brown-John, M., Chand, S., Featherstone, R., Masson, A.,  
 Mayo, M., Moran, J., Olson, T., Wong, D., Friedmann, M.F., Ritland, C.E.,  
 Siddiqui, A., Holt, R., Jones, S., Marra, M., Ellis, B.E., Douglas, C.,  
 Ritland, K. and Bohlmann, J.

The spruce transcriptome: Analysis of expressed sequence tags from  
 multiple cDNA libraries

## JOURNAL

## COMMENT

Contact: Joerg Bohlmann  
 Genome BC forest genomics program  
 University of British Columbia  
 Michael Smith Laboratories, 6174 University Boulevard, Rm. 237,  
 Vancouver, British Columbia, Canada, V6T 1Z3  
 Tel: 1-604-822-0282  
 Fax: 1-604-822-2114

## Email: bohlmann@msl.ubc.ca

## Plate: WS00962 row: M column: 09

## High quality sequence stop: 791

## POLYA=Yes.

## FEATURES

## Location/Qualifiers

## source

1. .791  
 /organism="Picea engelmannii x Picea sitchensis"  
 /mol\_type="mRNA"  
 /cultivar="Pal-1028"  
 /db\_xref="taxon:273280"  
 /clone="WS00962\_M09"  
 /sex="Hermaphrodite"  
 /lab\_host="E. coli DH10B cells"  
 /clone\_lib="IS-B-N-A-10"  
 /note="Organ: Bark (with phloem and cambium attached) from  
 one year old clonal trees grown under greenhouse  
 conditions in standard potting soil mixture; Vector:  
 pBluescript II SK (+) XR; Site 1: EcoRI (5' end of cDNA);  
 Site 2: XhoI (3' end of cDNA); Bark was wounded using  
 razor blades along the entire length of the tree at 5 mm  
 intervals on opposite sides of the trunk. The same trees  
 were also sprayed with a 0.01% (v/v) methyl jasmonate  
 solution resuspended in 0.1% (v/v) tween 20 (~50mLs per  
 tree). Bark tissue with phloem attached was harvested 3  
 hours, 6 hours, 12 hours, 24 hours, 2 days, 4 days and 8  
 days after initiating the treatment. Untreated control  
 bark was also harvested at time 0 hours. mRNA was isolated  
 from each tissue source independently and equal quantities  
 of mRNA from each tissue were then pooled. cDNA was  
 prepared from 5 micrograms of mRNA and directionally  
 ligated into the pBluescript II SK (+) XR vector using the  
 pBluescript II XR cDNA Library Construction kit according  
 to manufacturer's instructions with modifications  
 (Stratagene). Plasmid DNA was then transformed by  
 electroporation into DH10B cells (Invitrogen) for  
 propagation. Normalization was applied according to



published methods [Bonaldo M.F. et al. (1996) Genome Research 6(9):791] in order to reduce the abundance of highly expressed transcripts."

## ORIGIN

Query Match 87.0%; Score 17.4; DB 8; Length 791;  
Best Local Similarity 94.7%; Pred. No. 2.4e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGCCAAACAGGACCAAC 20

|||||  
Db 213 AGCCAAACAGGACCAAC 231

## RESULT 34

BF791575

LOCUS 908 bp mRNA linear EST 12-JAN-2001  
DEFINITION 60251677F1 NIH\_MGC\_84 Homo sapiens cDNA clone IMAGE:4343922 5',  
mRNA sequence.

## ACCESSION

BF791575

BF791575.1 GI:12096629

## KEYWORDS

EST.

## SOURCE

Homo sapiens

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

## REFERENCE

1 (bases 1 to 908)

## AUTHORS

NIH-MGC

http://mgs.nci.nih.gov/

## TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

## JOURNAL

Unpublished (1999)

## COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgabbs@mail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LAM961 row: j column: 19

High quality sequence stop: 631.

Location/Qualifiers

1..908

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:4343922"

/tissue\_type="adrenal cortex carcinoma, cell line"

/lab\_host="DH10B (phage-resistant)"

/clone\_lib="NIH MGC 84"

/notes="Organ: adrenal gland; Vector: pCMV-SPORT6; Site\_1:

NotI; Site\_2: SalI; Cloned unidirectionally; oligo-dT:

primed. Average insert size 1.229 Kb. Library enriched for

full-length clones and constructed by Life Technologies.

Note: this is a NIH\_MGC library."

ORIGIN

Query Match 87.0%; Score 17.4; DB 2; Length 908;

Best Local Similarity 94.7%; Pred. No. 2.4e+03;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGACCAAA 19

|||||

Db 890 GAGCCAAACAGGACCAAA 908

|||||

RESULT 35

CNS020RK/c

LOCUS

DEFINITION

## ACCESSION

AL175961

## VERSION

GI:7814018

## KEYWORDS

GSS; genome survey sequence.

## SOURCE

Tetraodon nigroviridis

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;

Tetraodontidae; Tetraodontidae; Tetraodon.

## REFERENCE

1

## AUTHORS

Roest Crolius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,

Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,

Saurin, W., Bernot, A., and Weissenbach, J.

## TITLE

Estimate of human gene number provided by genome-wide analysis

using Tetraodon nigroviridis DNA sequence

Nat. Genet. 25 (2), 235-238 (2000)

## JOURNAL

PUBMED

10835645

## REFERENCE

2

## AUTHORS

Roest Crolius, H., Jaillon, O., Dasilva, C., Ozouf-Costaz, C.,

Fizames, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F.,

Saurin, W., Bernot, A., and Weissenbach, J.

## TITLE

Characterization and repeat analysis of the compact genome of the

freshwater pufferfish Tetraodon nigroviridis

Genome Res. 10 (7), 939-949 (2000)

## JOURNAL

PUBMED

10899143

## REFERENCE

3 (bases 1 to 914)

## AUTHORS

Direct Submission

## TITLE

Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

## JOURNAL

COMMENT

This sequence is a single read and was generated as part of a large

scale clone-end sequencing project of the Tetraodon nigroviridis

genome. For more information, please take a look at

http://www.genoscope.cns.fr/Tetraodon.

## FEATURES

source

1..914

/organism="Tetraodon nigroviridis"

/mol\_type="genomic DNA"

/db\_xref="taxon:99883"

/clone="224G17"

/clone\_lib="G"

/note="Genoscope sequence ID : C0AG224AD09SP1

end : PUC-Ori"

## ORIGIN

Query Match 87.0%; Score 17.4; DB 10; Length 914;

Best Local Similarity 94.7%; Pred. No. 2.4e+03;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGCCAAACAGGACCAAC 20

|||||

Db 558 AGCCAAACAGGACCAAC 540

|||||

RESULT 36

CNS03MM7

LOCUS

DEFINITION

Tetraodon nigroviridis genome survey sequence R7 end of clone

038M17 of library G from Tetraodon nigroviridis, genomic survey

sequence.

AL250936

AL250936.1 GI:7971948

GSS; genome survey sequence.

Tetraodon nigroviridis

Tetraodon nigroviridis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;

Tetraodontidae; Tetraodontidae; Tetraodon.

1

Roest Crolius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,

Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

Saurin, W., Bernot, A., and Weissenbach, J.

```

Saurin,W. and Weissenbach,J.
Estimate of human gene number provided by genome-wide analysis
using Tetraodon nigroviridis DNA sequence
Nat. Genet. 25 (2), 235-238 (2000)
10835645
2
Roest Crolius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
Saurin,W., Bernot,A. and Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Genome Res. 10 (7), 939-949 (2000)
10839143
3 (bases 1 to 933)
Direct Submission
Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
Location/Qualifiers
1..933
/organism="Tetraodon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone="038M17"
/clone_lib="G"
/note="Genoscope sequence ID : COBG038AG09LP1
end : T7"

ORIGIN
Query Match 87.0%; Score 17.4; DB 11; Length 933;
Best Local Similarity 94.7%; Pred. No. 2.4e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGCCAAACAGGAACCAAC 20
|||||
DB 49 AGCCAAACAGGAACCAAC 67

RESULT 37
CG859874 1012 bp DNA linear GSS 19-NOV-2003
LOCUS ZMMBBc0269F16r ZMMBBc (EcoRI) Zea mays genomic clone ZMMBBc0269F16
DEFINITION 3', genomic survey sequence.
ACCESSION CG859874
VERSION CG859874.1 GI:38432587
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 1012)
Zohovetz,V., Fuks,G., Yu,Y., Wing,R. and Messing,J.
Sequencing of the maize genome at PGIR (2003c)
Unpublished (2003)
Contact: Bharti,A.K.
Dr.Joachim Messing's lab
The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers
University
190 Frelinghuysen Road, Piscataway, NJ 08854, USA
Tel: 732 445 3801
Fax: 732 445 5735
Email: Bharti@waksman.rutgers.edu
Seq primer: SP6
Class: BAC ends
High quality sequence start: 87.
Location/Qualifiers

```

```

1..1012
/organism="Zea mays"
/mol_type="genomic DNA"
/cultivar="B73"
/db_xref="taxon:4577"
/db_xref="taxon:4577"
/clone="ZMMBBc0269F16"
/lab_host="E. coli DH10B"
/clone_lib="ZMMBBc (EcoRI)"
/note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI"

ORIGIN
Query Match 87.0%; Score 17.4; DB 10; Length 1012;
Best Local Similarity 94.7%; Pred. No. 2.4e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGCCAAACAGGAACCAAC 20
|||||
DB 907 AGCCAAACAGGAATCAAC 925

RESULT 38
CNS05FQU 1016 bp DNA linear GSS 01-SEP-2000
LOCUS Tetraodon nigroviridis genome survey sequence T3 end of clone
DEFINITION 004G06 of library A from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL335343
VERSION AL335343.1 GI:8229101
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontoidea; Tetraodontidae; Tetraodon.
1
Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Estimate of human gene number provided by genome-wide analysis
using Tetraodon nigroviridis DNA sequence
Nat. Genet. 25 (2), 235-238 (2000)
10835645
2
Roest Crolius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
Saurin,W., Bernot,A. and Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Genome Res. 10 (7), 939-949 (2000)
10839143
3 (bases 1 to 1016)
Direct Submission
Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
Location/Qualifiers
1..1016
/organism="Tetraodon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone="004G06"
/clone_lib="A"
/note="Genoscope sequence ID : C0AA004BD03A1
end : T3"

ORIGIN
Query Match 87.0%; Score 17.4; DB 11; Length 1016;

```

```
Best Local Similarity 94.7%; Pred. No. 2.4e+03; Mismatches 1; Indels 0; Gaps 0;
Matches 18; Conservative 0;

QY 2 AGCCAAACAGGAACCAAC 20
    |||||
Db 699 AGCCCAACAGGAACCAAC 717

RESULT 39
CW923704/c
LOCUS
DEFINITION
    CW923704 1022 bp DNA linear GSS 24-FEB-2005
    EDCB010TR A. castellanii, 6-8 kb library from total genomic DNA
    Acanthamoeba castellanii genomic clone EDCB010, genomic survey
    sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
    ORGANISM
        Acanthamoeba castellanii
        Acanthamoeba castellanii
        Eukaryota; Acanthamoebidae; Acanthamoeba.
REFERENCE
    1 (bases 1 to 1022)
    Anderson, I.J. and Loftus, B.J.
    Gene discovery in the Acanthamoeba castellanii genome
    Unpublished (2004)
    CONTACT: Iain Anderson
    The Institute for Genomic Research (TIGR; www.tigr.org)
    9712 Medical Center Drive, Rockville, MD 20850, USA
    Tel: 301-795-7949
    Fax: 301-838-0208
    Class: shotgun.
FEATURES
    source
    1..1022
        /organism="Acanthamoeba castellanii"
        /mol_type="genomic DNA"
        /strain="Neff"
        /db_xref="taxon:5755"
        /clone="EDCB010"
        /clone_lib="A. castellanii, 6-8 kb library from total
        genomic DNA"
        /note="Vector: pHOS2"
ORIGIN
Query Match 87.0%; Score 17.4; DB 10; Length 1022;
Best Local Similarity 94.7%; Pred. No. 2.4e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAAA 19
    |||||
Db 321 GAGCCAAACAGGAACCAAA 303

RESULT 40
CW935554
LOCUS
DEFINITION
    CW935554 1119 bp DNA linear GSS 24-FEB-2005
    EDCCU52TR A. castellanii, 6-8 kb library from total genomic DNA
    Acanthamoeba castellanii genomic clone EDCCU52, genomic survey
    sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
    ORGANISM
        Acanthamoeba castellanii
        Acanthamoeba castellanii
        Eukaryota; Acanthamoebidae; Acanthamoeba.
REFERENCE
    1 (bases 1 to 1119)
    Anderson, I.J. and Loftus, B.J.
    Gene discovery in the Acanthamoeba castellanii genome
    Unpublished (2004)
    CONTACT: Iain Anderson
    The Institute for Genomic Research (TIGR; www.tigr.org)
    9712 Medical Center Drive, Rockville, MD 20850, USA
    Tel: 301-795-7949
    Fax: 301-838-0208
    Class: shotgun.
FEATURES
    source
    1..1119
        /organism="Acanthamoeba castellanii"
        /mol_type="genomic DNA"
        /strain="Neff"
        /db_xref="taxon:5755"
        /clone="EDCCU52"
        /clone_lib="A. castellanii, 6-8 kb library from total
        genomic DNA"
        /note="Vector: pHOS2"
ORIGIN
Query Match 87.0%; Score 17.4; DB 10; Length 1119;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAAA 19
    |||||
Db 1011 GAGCCAAACAGGAACCAAA 1029

Search completed: November 20, 2005, 21:55:32
Job time : 2994.11 secs
```

This Page Blank (uspto)

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	17.4	87.0	601	3	US-09-949-016-69821	Sequence 69821, A
C 2	17.4	87.0	39318	3	US-09-949-016-113798	Sequence 113798, A
	17	85.0	248968	3	US-09-949-016-12614	Sequence 12614, A
C 3	17	85.0	250958	3	US-09-949-016-16061	Sequence 16061, A
C 4	16.8	84.0	2780	3	US-09-023-655-11339	Sequence 1139, A
C 5	16.8	84.0	12122	3	US-09-949-016-16902	Sequence 16902, A
C 6	16.8	84.0	199945	3	US-09-949-016-15436	Sequence 15436, A
C 7	16.8	84.0	62386	3	US-09-949-016-12823	Sequence 12823, A
C 8	16.4	82.0	62386	3	US-09-949-016-14546	Sequence 14546, A
C 9	16.4	82.0	818128	3	US-09-949-016-14547	Sequence 14547, A
C 10	16.4	82.0	818128	3	US-09-949-016-14548	Sequence 14548, A
C 11	16.4	82.0	818128	3	US-09-949-016-14549	Sequence 14549, A
C 12	16.4	82.0	818128	3	US-09-949-016-14550	Sequence 14550, A
C 13	16.4	82.0	818128	3	US-09-949-016-14551	Sequence 14551, A
C 14	16.4	82.0	818128	3	US-09-949-016-14552	Sequence 14552, A
C 15	16.4	82.0	818128	3	US-09-949-016-14553	Sequence 14553, A
C 16	16.4	82.0	818128	3	US-09-949-016-14554	Sequence 14554, A
C 17	16.4	82.0	818128	3	US-09-949-016-14555	Sequence 14555, A
C 18	16.4	82.0	818128	3	US-09-949-016-14556	Sequence 14556, A
C 19	16.4	82.0	818128	3	US-09-949-016-14557	Sequence 14557, A
C 20	16.4	82.0	818128	3	US-09-949-016-14558	Sequence 14558, A
C 21	16.4	82.0	818128	3	US-09-949-016-14559	Sequence 14559, A
C 22	16.4	82.0	818128	3	US-09-949-016-14560	Sequence 14560, A
C 23	16.4	82.0	818128	3	US-09-949-016-14561	Sequence 14561, A

; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13798  
; LENGTH: 39318  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-13798

Query Match 87.0%; Score 17.4; DB 3; Length 39318;  
Best Local Similarity 94.7%; Pred. No. 1.1e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAA 19  
||||| ||||| ||||| ||||| |||||  
Db 27487 GAGCCCAACAGGAACCAA 27469

RESULT 3  
US-09-949-016-12614/c  
; Sequence 12614, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12614  
; LENGTH: 248968  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(248968)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-12614

Query Match 85.0%; Score 17; DB 3; Length 248968;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AGCCAAACAGGAACCAA 18  
||||| ||||| ||||| ||||| |||||  
Db 148788 AGCCAAACAGGAACCAA 148772

RESULT 4  
US-09-949-016-16061/c  
; Sequence 16061, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16061  
; LENGTH: 250958  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(250958)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-16061

Query Match 85.0%; Score 17; DB 3; Length 250958;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AGCCAAACAGGAACCAA 18  
||||| ||||| ||||| ||||| |||||  
Db 198778 AGCCAAACAGGAACCAA 198762

RESULT 5  
US-09-023-655-1139/c  
; Sequence 1139, Application US/09023655  
; Patent No. 6607879  
; GENERAL INFORMATION:  
; APPLICANT: Cocks, Benjamin G.  
; APPLICANT: Susan G. Stuart  
; APPLICANT: Jeffrey J. Seilhamer  
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE  
; TITLE OF INVENTION: EXPRESSION  
; NUMBER OF SEQUENCES: 1508  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/023,655  
; FILING DATE: HERewith  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Zeller, Karen J.  
; REGISTRATION NUMBER: 37,071  
; REFERENCE/DOCKET NUMBER: PA-0001 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 855-0555  
; TELEFAX: (650) 845-4166  
; INFORMATION FOR SEQ ID NO: 1139:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2780 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: GENBANK  
; CLONE: g186508  
US-09-023-655-1139

```
Query Match      84.0%; Score 16.8; DB 3; Length 2780;
Best Local Similarity 90.0%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 GAGCCAAACAGGAACCAAAAC 20
      ||||| ||||| ||||| ||||| |||||
Db      122 GAGCCAAACAGGAACAGAC 103

RESULT 6
US-09-949-016-16902/c
; Sequence 16902, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16902
; LENGTH: 12122
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16902

Query Match      84.0%; Score 16.8; DB 3; Length 12122;
Best Local Similarity 90.0%; Pred. No. 1.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 GAGCCAAACAGGAACCAAAAC 20
      ||||| ||||| ||||| ||||| |||||
Db      1349 GAGCCAAACAGGAACAGAC 1330

RESULT 7
US-09-949-016-15436/c
; Sequence 15436, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15436
; LENGTH: 199945
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15436

Query Match      84.0%; Score 16.8; DB 3; Length 199945;
Best Local Similarity 90.0%; Pred. No. 2.6e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 GAGCCAAACAGGAACCAAAAC 20
      ||||| ||||| ||||| ||||| |||||
Db      189530 GAGAGAAACAGGAACCAAAAC 189511

RESULT 8
US-09-949-016-12823
; Sequence 12823, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12823
; LENGTH: 62386
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc feature
; LOCATION: (1)...(62386)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12823

Query Match      82.0%; Score 16.4; DB 3; Length 62386;
Best Local Similarity 94.4%; Pred. No. 3.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 AGCCAAACAGGAACCAAA 19
      ||||| ||||| ||||| ||||| |||||
Db      43186 AGCTAAACAGGAACCAAA 43203

RESULT 9
US-09-949-016-14546/c
; Sequence 14546, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14546
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14546

Query Match      82.0%; Score 16.4; DB 3; Length 818128;
Best Local Similarity 94.4%; Pred. No. 4.5e+02;
```

```
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 AGCCAAACAGGACCAAA 19
    ||||| ||||| ||||| |||||
Db 417013 AGCCATACAGGAACCAA 416996

RESULT 10
US-09-949-016-14547/c
; Sequence 14547, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14547
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14547

Query Match 82.0%; Score 16.4; DB 3; Length 818128;
Best Local Similarity 94.4%; Pred. No. 4.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 AGCCAAACAGGACCAAA 19
    ||||| ||||| ||||| |||||
Db 417013 AGCCATACAGGACCAAA 416996

RESULT 11
US-09-949-016-14548/c
; Sequence 14548, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14548
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14548

Query Match 82.0%; Score 16.4; DB 3; Length 818128;
Best Local Similarity 94.4%; Pred. No. 4.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 AGCCAAACAGGACCAAA 19
    ||||| ||||| ||||| |||||
Db 417013 AGCCATACAGGACCAAA 416996

RESULT 12
US-09-949-016-14549/c
; Sequence 14549, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14549
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14549

Query Match 82.0%; Score 16.4; DB 3; Length 818128;
Best Local Similarity 94.4%; Pred. No. 4.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 AGCCAAACAGGACCAAA 19
    ||||| ||||| ||||| |||||
Db 417013 AGCCATACAGGACCAAA 416996

RESULT 13
US-09-949-016-14550/c
; Sequence 14550, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14550
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc_feature
```

```
Query Match 82.0%; Score 16.4; DB 3; Length 818128;
Best Local Similarity 94.4%; Pred. No. 4.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 AGCCAAACAGGACCAAA 19
    ||||| ||||| ||||| |||||
Db 417013 AGCCATACAGGACCAAA 416996

RESULT 12
US-09-949-016-14549/c
; Sequence 14549, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14549
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14549

Query Match 82.0%; Score 16.4; DB 3; Length 818128;
Best Local Similarity 94.4%; Pred. No. 4.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 AGCCAAACAGGACCAAA 19
    ||||| ||||| ||||| |||||
Db 417013 AGCCATACAGGACCAAA 416996

RESULT 13
US-09-949-016-14550/c
; Sequence 14550, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14550
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc_feature
```



```
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14550

Query Match      82.0%; Score 16.4; DB 3; Length 818128;
Best Local Similarity 94.4%; Pred. No. 4.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGCCAAACAGGAACCAAA 19
Db 417013 AGCCATACAGGAACCAAA 416996

RESULT 14
US-09-949-016-14551/c
; Sequence 14551, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14551
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14551

Query Match      82.0%; Score 16.4; DB 3; Length 818128;
Best Local Similarity 94.4%; Pred. No. 4.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGCCAAACAGGAACCAAA 19
Db 417013 AGCCATACAGGAACCAAA 416996

RESULT 15
US-09-949-016-14552/c
; Sequence 14552, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14552
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14552

Query Match      82.0%; Score 16.4; DB 3; Length 818128;
Best Local Similarity 94.4%; Pred. No. 4.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGCCAAACAGGAACCAAA 19
Db 417013 AGCCATACAGGAACCAAA 416996

RESULT 16
US-09-949-016-14553/c
; Sequence 14553, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14553
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14553

Query Match      82.0%; Score 16.4; DB 3; Length 818128;
Best Local Similarity 94.4%; Pred. No. 4.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGCCAAACAGGAACCAAA 19
Db 417013 AGCCATACAGGAACCAAA 416996

RESULT 17
US-09-949-016-14554/c
; Sequence 14554, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14554
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14554
```

```
; SEQ ID NO 14554
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14554

Query Match      82.0%; Score 16.4; DB 3; Length 818128;
Best Local Similarity 94.4%; Pred. No. 4.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGCCAAACAGGAACCAAA 19
Db 417013 AGCCATACAGGAACCAAA 416996

RESULT 18
US-09-949-016-14555/c
; Sequence 14555, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14555
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14555

Query Match      82.0%; Score 16.4; DB 3; Length 818128;
Best Local Similarity 94.4%; Pred. No. 4.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGCCAAACAGGAACCAAA 19
Db 417013 AGCCATACAGGAACCAAA 416996

RESULT 19
US-09-949-016-14556/c
; Sequence 14556, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14556
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14556

Query Match      82.0%; Score 16.4; DB 3; Length 818128;
Best Local Similarity 94.4%; Pred. No. 4.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGCCAAACAGGAACCAAA 19
Db 417013 AGCCATACAGGAACCAAA 416996

RESULT 20
US-09-949-016-14557/c
; Sequence 14557, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14557
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14557

Query Match      82.0%; Score 16.4; DB 3; Length 818128;
Best Local Similarity 94.4%; Pred. No. 4.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGCCAAACAGGAACCAAA 19
Db 417013 AGCCATACAGGAACCAAA 416996

RESULT 21
US-09-949-016-14558/c
; Sequence 14558, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14558
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14558

Query Match      82.0%; Score 16.4; DB 3; Length 818128;
Best Local Similarity 94.4%; Pred. No. 4.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGCCAAACAGGAACCAAA 19
Db 417013 AGCCATACAGGAACCAAA 416996
```

; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14558  
; LENGTH: 818128  
; TYPE: DNA  
; ORGANISM: Human  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(818128)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-14558

Query Match 82.0%; Score 16.4; DB 3; Length 818128;  
Best Local Similarity 94.4%; Pred. No. 4.5e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AGCCAAACAGGAACCAAA 19  
Db 417013 AGCCATACAGGAACCAAA 416996  
||||| |||||||

## RESULT 22

US-09-949-016-14559/c  
; Sequence 14559, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14559  
; LENGTH: 818128  
; TYPE: DNA  
; ORGANISM: Human  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(818128)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-14559

Query Match 82.0%; Score 16.4; DB 3; Length 818128;  
Best Local Similarity 94.4%; Pred. No. 4.5e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AGCCAAACAGGAACCAAA 19  
Db 417013 AGCCATACAGGAACCAAA 416996  
||||| |||||||

## RESULT 23

US-09-949-016-14560/c  
; Sequence 14560, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14560  
; LENGTH: 818128  
; TYPE: DNA  
; ORGANISM: Human  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(818128)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-14560

Query Match 82.0%; Score 16.4; DB 3; Length 818128;  
Best Local Similarity 94.4%; Pred. No. 4.5e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AGCCAAACAGGAACCAAA 19  
Db 417013 AGCCATACAGGAACCAAA 416996  
||||| |||||||

## RESULT 24

US-09-949-016-14561/c  
; Sequence 14561, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14561  
; LENGTH: 818128  
; TYPE: DNA  
; ORGANISM: Human  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(818128)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-14561

Query Match 82.0%; Score 16.4; DB 3; Length 818128;  
Best Local Similarity 94.4%; Pred. No. 4.5e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AGCCAAACAGGAACCAAA 19  
Db 417013 AGCCATACAGGAACCAAA 416996  
||||| |||||||

## RESULT 25

US-09-949-016-14562/c  
; Sequence 14562, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

```
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14562
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14562

Query Match      82.0%; Score 16.4; DB 3; Length 818128;
Best Local Similarity 94.4%; Pred. No. 4.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGCCAAACAGGACCAAA 19
Db 417013 AGCCATACAGGAACCAAA 416996

RESULT 26
US-09-949-016-14564/c
; Sequence 14564, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14564
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14564

Query Match      82.0%; Score 16.4; DB 3; Length 818128;
Best Local Similarity 94.4%; Pred. No. 4.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGCCAAACAGGACCAAA 19
Db 417013 AGCCATACAGGAACCAAA 416996

RESULT 27
US-09-949-016-14565/c
; Sequence 14565, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14565
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14565

Query Match      82.0%; Score 16.4; DB 3; Length 818128;
Best Local Similarity 94.4%; Pred. No. 4.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGCCAAACAGGACCAAA 19
Db 417013 AGCCATACAGGAACCAAA 416996

RESULT 28
US-09-949-016-14566/c
; Sequence 14566, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14566
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14566

Query Match      82.0%; Score 16.4; DB 3; Length 818128;
Best Local Similarity 94.4%; Pred. No. 4.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGCCAAACAGGACCAAA 19
Db 417013 AGCCATACAGGAACCAAA 416996

RESULT 29
```

```
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14565
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14565

Query Match      82.0%; Score 16.4; DB 3; Length 818128;
Best Local Similarity 94.4%; Pred. No. 4.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGCCAAACAGGACCAAA 19
Db 417013 AGCCATACAGGAACCAAA 416996

RESULT 28
US-09-949-016-14566/c
; Sequence 14566, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14566
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14566

Query Match      82.0%; Score 16.4; DB 3; Length 818128;
Best Local Similarity 94.4%; Pred. No. 4.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGCCAAACAGGACCAAA 19
Db 417013 AGCCATACAGGAACCAAA 416996

RESULT 29
```

US-09-949-016-14567/c  
; Sequence 14567, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14567  
; LENGTH: 818128  
; TYPE: DNA  
; ORGANISM: Human  
; NAME/KEY: misc feature  
; LOCATION: (1)...(818128)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-14567

Query Match 82.0%; Score 16.4; DB 3; Length 818128;  
Best Local Similarity 94.4%; Pred. No. 4.5e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGCCAAACAGGAACCAAA 19  
||||| ||||| ||||| ||||| |||||  
DB 417013 AGCCATACAGGAACCAAA 416996

RESULT 30  
US-08-852-807-11/c  
; Sequence 11, Application US/08852807  
; Patent No. 5861298  
; GENERAL INFORMATION:  
; APPLICANT: Debouck, Christine  
; APPLICANT: Drake, Fred  
; APPLICANT: Gowen, Maxine  
; APPLICANT: Rood, Julie  
; APPLICANT: Hastings, Gregg  
; APPLICANT: Adams, Mark  
; APPLICANT: Fraser, Claire  
; APPLICANT: Lee, No. 5861298man  
; APPLICANT: Kirkness, Ewen  
; APPLICANT: Blake, Judith  
; APPLICANT: Fitzgerald, Lisa  
; TITLE OF INVENTION: CATHEPSIN K GENE  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SmithKline Beecham Corporation  
; STREET: 709 Swedeland Road  
; CITY: King of Prussia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19406-2799  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/852,807  
; FILING DATE: 07-MAY-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/019,942

; FILING DATE: 14-JUNE-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/020,273  
; FILING DATE: 17-JUNE-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/026,273  
; FILING DATE: 26-AUG-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Han, William T  
; REGISTRATION NUMBER: 34,344  
; REFERENCE/DOCKET NUMBER: ATG50006-2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 610-270-5219  
; TELEFAX: 610-270-5090  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 219 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE:  
; ORIGINAL SOURCE:  
US-08-852-807-11

Query Match 79.0%; Score 15.8; DB 2; Length 219;  
Best Local Similarity 89.5%; Pred. No. 3e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGCCAAACAGGAACCAAC 20  
||||| ||||| ||||| ||||| |||||  
DB 24 AGCCCAACAGGAACCCAC 6

RESULT 31  
US-09-949-016-38380/c  
; Sequence 38380, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 38380  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-38380

Query Match 79.0%; Score 15.8; DB 3; Length 601;  
Best Local Similarity 89.5%; Pred. No. 3.5e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAAA 19  
||||| ||||| ||||| ||||| |||||  
DB 593 GAGCAACAGGAACCAAA 575

RESULT 32  
US-09-949-016-56963/c

```
; Sequence 56963, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 56963
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-56963

Query Match          79.0%; Score 15.8; DB 3; Length 601;
Best Local Similarity 89.5%; Pred. No. 3.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAA 19
    ||||| ||||| ||||| |||||
Db 445 GAGCAAAACTGGAACCAA 427

RESULT 33
US-09-949-016-201233/c
; Sequence 201233, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 201233
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-201233

Query Match          79.0%; Score 15.8; DB 3; Length 601;
Best Local Similarity 89.5%; Pred. No. 3.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAA 19
    ||||| ||||| ||||| |||||
Db 593 GAGCAAACTGGAACCAA 575

RESULT 34
US-08-592-214A-5
; Sequence 5, Application US/08592214A
; Patent No. 5811536
; GENERAL INFORMATION:
; APPLICANT: Yanofsky, Martin F.
; TITLE OF INVENTION: Cauliflower Floral Meristem Identify
```

```
; TITLE OF INVENTION: Genes and Methods of Using Same
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,214A
; FILING DATE: 26-JAN-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-UD 1927
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 768 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..766
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..768
; OTHER INFORMATION: /note= "product = Brassica oleracea
; var. botrytis API"
US-08-592-214A-5

Query Match          79.0%; Score 15.8; DB 2; Length 768;
Best Local Similarity 89.5%; Pred. No. 3.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGCCAAACAGGAACCAAAAC 20
    ||||| ||||| ||||| |||||
Db 477 AGCCATACAGGAACCAAAAC 495

RESULT 35
US-08-659-188-5
; Sequence 5, Application US/08659188
; Patent No. 6002069
; GENERAL INFORMATION:
; APPLICANT: Yanofsky, Martin F.
; TITLE OF INVENTION: Seed Plants Exhibiting Inducible Early
; Reproductive Development and Methods of Making Same
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/08/659,188
; FILING DATE: 05-JUN-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-UD 1946
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 768 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..766
; NAME/KEY: misc feature
; LOCATION: 1..768
; OTHER INFORMATION: /note= "product = Brassica oleracea
; OTHER INFORMATION: var. botrytis AP1."
; US-08-659-188-5

Query Match 79.0%; Score 15.8; DB 3; Length 768;
Best Local Similarity 89.5%; Pred. No. 3.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGCCAAACAGGAACCAAAAC 20
    ||||| ||||| ||||| |||||
Db 477 AGCCATACAGGAACAAAC 495

RESULT 36
US-08-655-227-5
; Sequence 5, Application US/08655227
; Patent No. 6025483
; GENERAL INFORMATION:
; APPLICANT: Yanofsky, Martin F.
; TITLE OF INVENTION: Maize and Cauliflower APETALAL Gene
; TITLE OF INVENTION: Products and Nucleic Acid Molecules Encoding Same
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/655,227
; FILING DATE: 05-JUN-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-UD 2143
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 768 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
```

```
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..766
; NAME/KEY: misc feature
; LOCATION: 1..768
; OTHER INFORMATION: /note= "product = Brassica oleracea
; OTHER INFORMATION: var. botrytis AP1."
; US-08-655-227-5

Query Match 79.0%; Score 15.8; DB 3; Length 768;
Best Local Similarity 89.5%; Pred. No. 3.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGCCAAACAGGAACCAAAAC 20
    ||||| ||||| ||||| |||||
Db 477 AGCCATACAGGAACAAAC 495

RESULT 37
US-08-655-241-5
; Sequence 5, Application US/08655241
; Patent No. 6025543
; GENERAL INFORMATION:
; APPLICANT: Yanofsky, Martin F.
; APPLICANT: Weigel, Detlef
; TITLE OF INVENTION: Seed Plants Exhibiting Early Reproductive
; TITLE OF INVENTION: Development and Methods of Making Same
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/655,241
; FILING DATE: 05-JUN-1996
; CLASSIFICATION: CLASS 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-UD 1894
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 768 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..766
; NAME/KEY: misc feature
; LOCATION: 1..768
; OTHER INFORMATION: /note= "product = Brassica oleracea
; OTHER INFORMATION: var. botrytis AP1."
; US-08-655-241-5

Query Match 79.0%; Score 15.8; DB 3; Length 768;
Best Local Similarity 89.5%; Pred. No. 3.6e+02;
```

```
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGCCAAACAGGAACCAAC 20
    ||||| ||||| |||||
Db 477 AGCCATACAGGAACAAAC 495

RESULT 38
US-09-149-976-5
; Sequence 5, Application US/09149976
; Patent No. 6127123
; GENERAL INFORMATION:
; APPLICANT: Yanofsky, Martin F.
; TITLE OF INVENTION: Cauliflower Floral Meristem Identity
; TITLE OF INVENTION: Genes and Methods of Using Same
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/149,976
; FILING DATE: 09-SEP-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/592,214
; FILING DATE: 26-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-UD 3291
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 768 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..766
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..768
; OTHER INFORMATION: /note= "product = Brassica oleracea
; OTHER INFORMATION: var. botrytis API"
US-09-149-976-5

Query Match 79.0%; Score 15.8; DB 3; Length 768;
Best Local Similarity 89.5%; Pred. No. 3.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGCCAAACAGGAACCAAC 20
    ||||| ||||| |||||
Db 477 AGCCATACAGGAACAAAC 495

RESULT 39
US-09-398-326-5
; Sequence 5, Application US/09398326
; Patent No. 6355863
; GENERAL INFORMATION:
```

```
; APPLICANT: Yanofsky, Martin F.
; TITLE OF INVENTION: Seed Plants Exhibiting Inducible Early
; TITLE OF INVENTION: Reproductive Development and Methods of Making Same
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/398,326
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/659,188
; FILING DATE: 05-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-UD 3739
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 768 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..766
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..768
; OTHER INFORMATION: /note= "product = Brassica oleracea
; OTHER INFORMATION: var. botrytis API."
US-09-398-326-5

Query Match 79.0%; Score 15.8; DB 3; Length 768;
Best Local Similarity 89.5%; Pred. No. 3.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGCCAAACAGGAACCAAC 20
    ||||| ||||| |||||
Db 477 AGCCATACAGGAACAAAC 495

RESULT 40
US-09-853-450-5
; Sequence 5, Application US/09853450
; Patent No. 6828478
; GENERAL INFORMATION:
; APPLICANT: Yanofsky, Martin F.
; APPLICANT: Pelaz, Soraya
; APPLICANT: Ditta, Gary
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Combinations of Genes for Producing Seed Plants
; TITLE OF INVENTION: Exhibiting Modulated Reproductive Development
; FILE REFERENCE: 19452A-002400US
; CURRENT APPLICATION NUMBER: US/09/853,450
; CURRENT FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
```



; LENGTH: 768  
; TYPE: DNA  
; ORGANISM: Brassica oleracea var. botrytis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(768)  
; OTHER INFORMATION: APETALA1 (AP1)  
US-09-853-450-5

Query Match 79.0% Score 15.8; DB 3; Length 768;  
Best Local Similarity 89.5%; Pred. No. 3.6e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AGCCAAACACGGAACCAAAC 20  
| | | | | | | | | | | | | | | | | |  
Db 477 AGCATACAGGAACAAAC 495

Search completed: November 20, 2005, 22:02:07  
Job time : 110.111 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2005, 18:34:25 ; Search time 602.469 Seconds  
(without alignments)  
274.516 Million cell updates/sec

Title: US-10-627-757-22

Perfect score: 20

Sequence: 1 gagccaaacaggaaacaaac 20

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_NA\_Main.\*  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*  
3: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*  
4: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*  
6: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*  
8: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/1/pubpna/US11\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	20	100.0	20	7	US-10-627-757-22
2	20	100.0	1074	7	US-10-627-757-5
3	20	100.0	46951	6	US-10-091-281-2
4	17.4	87.0	615	4	US-09-925-065A-35409
5	17.4	87.0	72732	7	US-10-052-482-193
6	17.4	87.0	410846	9	US-10-481-613-1
7	17	85.0	25	7	US-10-719-956-153198
8	17	85.0	263852	8	US-10-812-232-6
9	17	85.0	317129	8	US-10-741-600-17765
10	16.8	84.0	1843	8	US-10-425-115-62251
11	16.8	84.0	2742	6	US-10-173-551-37
12	16.8	84.0	2780	7	US-10-641-643-1139
13	16.8	84.0	2798	3	US-09-954-456-1145
14	16.8	84.0	2798	6	US-10-080-334-11
15	16.8	84.0	2798	9	US-10-843-641A-4172
16	16.8	84.0	2835	6	US-10-240-965-126
17	16.8	84.0	31766	3	US-09-765-344-5
18	16.8	84.0	31766	5	US-10-288-478-5
19	16.8	84.0	786452	8	US-10-719-993-6822
20	16.4	82.0	225	8	US-10-425-115-162723
21	16.4	82.0	545	4	US-09-925-065A-873201
22	16.4	82.0	570	7	US-10-424-599-81338
23	16.4	82.0	599	9	US-10-972-079-71375

C 24	16.4	82.0	600	9	US-10-972-079-42033	Sequence 42033, A
C 25	16.4	82.0	600	9	US-10-972-079-42034	Sequence 42034, A
C 26	16.4	82.0	600	9	US-10-972-079-42035	Sequence 42035, A
C 27	16.4	82.0	600	9	US-10-972-079-42036	Sequence 42036, A
C 28	16.4	82.0	600	9	US-10-972-079-42037	Sequence 42037, A
C 29	16.4	82.0	600	9	US-10-972-079-42038	Sequence 42038, A
C 30	16.4	82.0	745	7	US-10-424-599-74758	Sequence 74758, A
C 31	16.4	82.0	1195	7	US-10-424-599-80664	Sequence 80664, A
C 32	16.4	82.0	1557	4	US-09-925-065A-677481	Sequence 677481, A
C 33	16.4	82.0	1794	7	US-10-437-963-51993	Sequence 51993, A
C 34	16.4	82.0	2476	8	US-10-602-494-29	Sequence 29, Appl
C 35	16.4	82.0	2578	7	US-10-437-963-23160	Sequence 23160, A
C 36	16.4	82.0	301692	6	US-10-428-487-11	Sequence 11, Appl
C 37	16.4	82.0	310268	7	US-10-367-094-195	Sequence 195, App
C 38	16	80.0	707	7	US-10-767-701-13180	Sequence 13180, A
C 39	16	80.0	5735	10	US-11-097-143-32660	Sequence 32660, A
C 40	16	80.0	8954	10	US-11-097-143-32659	Sequence 32659, A
C 41	15.8	79.0	23	9	US-10-929-919A-9	Sequence 9, Appl
C 42	15.8	79.0	25	7	US-10-719-956-416556	Sequence 416556, A
C 43	15.8	79.0	25	10	US-11-036-317-52703	Sequence 52703, A
C 44	15.8	79.0	25	10	US-11-036-317-466740	Sequence 466740, A
C 45	15.8	79.0	25	10	US-11-036-317-788552	Sequence 788552, A

ALIGNMENTS

RESULT 1  
US-10-627-757-22  
; Sequence 22, Application US/10627757  
; Publication No. US20040091914A1  
; GENERAL INFORMATION:  
; APPLICANT: KOUCHI YASUHIRO  
; APPLICANT: MASASGO AKINORI  
; APPLICANT: TAKAHATI TAKAYUKI  
; TITLE OF INVENTION: GENE ASSAY METHOD FOR PREDICTING GLAUCOMA ONSET RISK  
; FILE REFERENCE: Q76319  
; CURRENT APPLICATION NUMBER: US/10/627,757  
; CURRENT FILING DATE: 2003-07-28  
; PRIOR APPLICATION NUMBER: JP P2002-226612  
; PRIOR FILING DATE: 2002-08-02  
; NUMBER OF SEQ ID NOS: 40  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 22  
; LENGTH: 20  
; TYPE: DNA  
; ORGANISM: Artificial  
; FEATURE:  
; OTHER INFORMATION: Designed DNA based on OPTN gene  
US-10-627-757-22

Query Match 100.0%; Score 20; DB 7; Length 20;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GAGCCAAACAGGAAACCAAC 20  
|||||  
Db 1 GAGCCAAACAGGAAACCAAC 20

RESULT 2  
US-10-627-757-5/c  
; Sequence 5, Application US/10627757  
; Publication No. US20040091914A1  
; GENERAL INFORMATION:  
; APPLICANT: KOUCHI YASUHIRO  
; APPLICANT: MASASGO AKINORI  
; APPLICANT: TAKAHATI TAKAYUKI  
; TITLE OF INVENTION: GENE ASSAY METHOD FOR PREDICTING GLAUCOMA ONSET RISK  
; FILE REFERENCE: Q76319  
; CURRENT APPLICATION NUMBER: US/10/627,757  
; CURRENT FILING DATE: 2003-07-28  
; PRIOR APPLICATION NUMBER: JP P2002-226612

```
; PRIOR FILING DATE: 2002-08-02
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 5
;   LENGTH: 1074
;   TYPE: DNA
;   ORGANISM: Homo sapiens
; US-10-627-757-5

Query Match      100.0%; Score 20; DB 7; Length 1074;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGACCAAAAC 20
   |||||
Db 784 GAGCCAAACAGGACCAAAAC 765

RESULT 3
US-10-091-281-2/c
; Sequence 2, Application US/10091281
; Publication No. US20030190617A1
; GENERAL INFORMATION:
; APPLICANT: RAYMOND, VINCENT
; APPLICANT: ST, ERWIN
; APPLICANT: MORISSETTE, JEAN
; TITLE OF INVENTION: OPTINEURIN NUCLEIC ACID MOLECULES AND USES THEREOF
; FILE REFERENCE: 13587.338
; CURRENT APPLICATION NUMBER: US/10/091,281
; CURRENT FILING DATE: 2002-03-06
; NUMBER OF SEQ ID NOS: 463
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 46951
;   TYPE: DNA
;   ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 391
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 691
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 709
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 887
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 894
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 987
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 1112
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 1505
; OTHER INFORMATION: insertion of additional c residue
; NAME/KEY: allele
; LOCATION: 1606
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 2405
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 2606
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 3313
; OTHER INFORMATION: single nucleotide polymorphism (SNP)

; NAME/KEY: allele
; LOCATION: 3555
; OTHER INFORMATION: insertion of additional t residue
; NAME/KEY: allele
; LOCATION: 3625
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 3629
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 3882
; OTHER INFORMATION: insertion of additional t residue
; NAME/KEY: allele
; LOCATION: 3988
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 4452
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: repeat region
; LOCATION: 598..878
; OTHER INFORMATION: repeat element
; NAME/KEY: repeat region
; LOCATION: 938..957
; OTHER INFORMATION: Short repeat element
; NAME/KEY: repeat region
; LOCATION: 1002..1329
; OTHER INFORMATION: ALU repeat element
; NAME/KEY: repeat region
; LOCATION: 2288..2587
; OTHER INFORMATION: ALU repeat element
; NAME/KEY: misc feature
; LOCATION: 5054_
; OTHER INFORMATION: putative transcription start site
; US-10-091-281-2

Query Match      100.0%; Score 20; DB 6; Length 46951;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGACCAAAAC 20
   |||||
Db 21394 GAGCCAAACAGGACCAAAAC 21375

RESULT 4
US-09-925-065A-35409
; Sequence 35409, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 35409
;   LENGTH: 615
;   TYPE: DNA
;   ORGANISM: Homo sapiens
; US-09-925-065A-35409
```

Query Match 87.0%; Score 17.4; DB 4; Length 615;  
Best Local Similarity 94.7%; Pred. No. 2.8e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAA 19  
||||| ||||| ||||| ||||| |||||  
Db 17 GAGCAAAACAGGAACCAA 35

RESULT 5  
US-10-052-482-193  
; Sequence 193, Application US/10052482  
; Publication No. US20040072264A1  
; GENERAL INFORMATION:  
; APPLICANT: Engelhard, Eric  
; APPLICANT: Morris, David  
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER  
; FILE REFERENCE: A-71087/RMS/DCF  
; CURRENT APPLICATION NUMBER: US/10/052,482  
; PRIOR FILING DATE: 2002-08-15  
; PRIOR APPLICATION NUMBER: US 09/747,377  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: US 09/798,586  
; PRIOR FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 241  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 193  
; LENGTH: 72732  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (3522)..(3541)  
; OTHER INFORMATION: "n" at positions 3522 to 3541 can be any base  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (19836)..(19911)  
; OTHER INFORMATION: "n" at positions 19836 to 19911 can be any base  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (43636)..(43770)  
; OTHER INFORMATION: "n" at positions 43636 to 43770 can be any base  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (47357)..(47376)  
; OTHER INFORMATION: "n" at positions 47357 to 47376 can be any base  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (57087)..(57120)  
; OTHER INFORMATION: "n" at positions 57087 to 57120 can be any base  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (64787)..(64806)  
; OTHER INFORMATION: "n" at positions 64787 to 64806 can be any base  
US-10-052-482-193

Query Match 87.0%; Score 17.4; DB 7; Length 72732;  
Best Local Similarity 94.7%; Pred. No. 3.3e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGCCAAACAGGAACCAAC 20  
||||| ||||| ||||| ||||| |||||  
Db 6897 AGCCAAACAGGAACCAAC 6915

RESULT 6  
US-10-481-613-1  
; Sequence 1, Application US/10481613  
; Publication No. US20050085627A1  
; GENERAL INFORMATION:  
; APPLICANT: Zhang, Youming  
; APPLICANT: Moffatt, Miriam  
; APPLICANT: Cookson, William

; APPLICANT: Tinsley, Jon  
; TITLE OF INVENTION: Atopy  
; FILE REFERENCE: 16721-0003US1 / P32688WO/KVC  
; CURRENT APPLICATION NUMBER: US/10/481,613  
; CURRENT FILING DATE: 2003-12-19  
; PRIOR APPLICATION NUMBER: PCT/GB02/02859  
; PRIOR FILING DATE: 2002-06-21  
; PRIOR APPLICATION NUMBER: GB 0115211.5  
; PRIOR FILING DATE: 2001-06-21  
; PRIOR APPLICATION NUMBER: GB 0115212.3  
; PRIOR FILING DATE: 2001-06-21  
; PRIOR APPLICATION NUMBER: GB 0115213.1  
; PRIOR FILING DATE: 2001-06-21  
; NUMBER OF SEQ ID NOS: 326  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 410846  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-481-613-1

Query Match 87.0%; Score 17.4; DB 9; Length 410846;  
Best Local Similarity 94.7%; Pred. No. 3.6e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAA 19  
||||| ||||| ||||| ||||| |||||  
Db 290923 GAGCAAAACAGGAACCAA 290941

RESULT 7  
US-10-719-956-153198/c  
; Sequence 153198, Application US/10719956  
; Publication No. US20040146910A1  
; GENERAL INFORMATION:  
; APPLICANT: Xue Mei Zhou  
; TITLE OF INVENTION: Methods of Genetic Analysis of Rat  
; FILE REFERENCE: 3527.1  
; CURRENT APPLICATION NUMBER: US/10/719,956  
; CURRENT FILING DATE: 2003-11-20  
; PRIOR APPLICATION NUMBER: 60/427,836  
; PRIOR FILING DATE: 2002-11-20  
; NUMBER OF SEQ ID NOS: 699466  
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1  
; SEQ ID NO 153198  
; LENGTH: 25  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
US-10-719-956-153198

Query Match 85.0%; Score 17; DB 7; Length 25;  
Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCA 17  
||||| ||||| ||||| ||||| |||||  
Db 25 GAGCCAAACAGGAACCA 9

RESULT 8  
US-10-812-232-6/c  
; Sequence 6, Application US/10812232  
; Publication No. US20040265961A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Wei  
; APPLICANT: Wu, Leeying  
; APPLICANT: Channavajhala, Padma L,  
; APPLICANT: Lin, Lih-hing  
; APPLICANT: Zhang, Yuhua  
; TITLE OF INVENTION: Novel Proteins Homologous to Kinase Suppressor of Ras  
; FILE REFERENCE: 01997.026700  
; CURRENT APPLICATION NUMBER: US/10/812,232  
; CURRENT FILING DATE: 2004-03-29

```
; PRIOR APPLICATION NUMBER: US 60/457,928
; PRIOR FILING DATE: 2003-03-28
; PRIOR APPLICATION NUMBER: US 60/491,283
; PRIOR FILING DATE: 2003-07-31
; NUMBER OF SEQ ID NOS: 254
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6
; LENGTH: 263852
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (3814)..(4319)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (6552)..(6571)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (17206)..(17225)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (19680)..(19699)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (24320)..(24339)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (26572)..(28591)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (32986)..(33005)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (56103)..(56724)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (59672)..(59691)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (61525)..(61544)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (71035)..(71054)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (79771)..(79807)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (91343)..(91801)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (141749)..(141768)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (144739)..(144758)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (164171)..(164340)
```

```
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (173689)..(173708)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (175051)..(175070)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (192924)..(192993)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (202010)..(202029)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (247837)..(247856)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (251852)..(252421)
; OTHER INFORMATION: n is a, c, g, or t
; US-10-812-232-6
```

```
Query Match      85.0%; Score 17; DB 8; Length 263852;
Best Local Similarity 100.0%; Pred. No. 5.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 GAGCCAAACAGGAACCA 17

DB 209107 GAGCCAAACAGGAACCA 209091

## RESULT 9

```
US-10-741-600-17765/c
; Sequence 17765, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001499
; CURRENT APPLICATION NUMBER: US/10/741,600
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 73997
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17765
; LENGTH: 317129
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)....(317129)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
; US-10-741-600-17765
```

```
Query Match      85.0%; Score 17; DB 8; Length 317129;
Best Local Similarity 100.0%; Pred. No. 5.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 2 AGCCAAACAGGAACCAA 18

DB 260941 AGCCAAACAGGAACCAA 260925

## RESULT 10

```
US-10-425-115-62251/c
; Sequence 62251, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
```







RESULT 16  
US-10-240-965-126/c  
; Sequence 126, Application US/10240965  
; Publication No. US20030165924A1  
; GENERAL INFORMATION:  
; APPLICANT: INCYTE GENOMICS, INC.  
; APPLICANT: SHIFFMAN, Dov  
; APPLICANT: SOMOGYI, Roland  
; APPLICANT: LAWN, Richard M.  
; APPLICANT: SEILHAMER, Jeffrey J.  
; APPLICANT: PORTER, Gordon J.  
; APPLICANT: MIKITA, Thomas  
; APPLICANT: TAL, Julie  
; TITLE OF INVENTION: GENES EXPRESSED IN FOAM CELL DIFFERENTIATION  
; FILE REFERENCE: PA-0025 PCT  
; CURRENT APPLICATION NUMBER: US/10/240,965  
; CURRENT FILING DATE: 2002-10-04  
; PRIOR APPLICATION NUMBER: 60/195,106  
; PRIOR FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 276  
; SOFTWARE: PERL Program  
; SEQ ID NO 126  
; LENGTH: 2835  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)...(31766)  
; OTHER INFORMATION: Incyte ID No. US20030165924A1 232066.3  
US-10-240-965-126

Query Match 84.0%; Score 16.8; DB 6; Length 2835;  
Best Local Similarity 90.0%; Pred. No. 5.6e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGACCAAC 20  
Db 162 GAGCCAAACAGGAAACAGAC 143

RESULT 17  
US-09-765-344-5/c  
; Sequence 5, Application US/09765344  
; Patent No. US20020019348A1  
; GENERAL INFORMATION:  
; APPLICANT: WEI, Ming-Hui et al  
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,  
; FILE REFERENCE: CL000746-CIP  
; CURRENT APPLICATION NUMBER: US/09/765,344  
; CURRENT FILING DATE: 2001-01-22  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 31766  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)...(31766)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-765-344-5

Query Match 84.0%; Score 16.8; DB 3; Length 31766;  
Best Local Similarity 90.0%; Pred. No. 6.2e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGACCAAC 20  
Db 11047 GAGACAAACAGGAAACAAAC 11028

RESULT 18  
US-10-288-478-5/c  
; Sequence 5, Application US/10288478  
; Publication No. US20030068784A1  
; GENERAL INFORMATION:  
; APPLICANT: WEI, Ming-Hui et al  
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,  
; FILE REFERENCE: CL000746-CON  
; CURRENT APPLICATION NUMBER: US/10/288,478  
; CURRENT FILING DATE: 2002-11-06  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 31766  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)...(31766)  
; OTHER INFORMATION: n = A,T,C or G  
US-10-288-478-5

Query Match 84.0%; Score 16.8; DB 5; Length 31766;  
Best Local Similarity 90.0%; Pred. No. 6.2e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGACCAAC 20  
Db 11047 GAGACAAACAGGAAACAAAC 11028

RESULT 19  
US-10-719-993-6822  
; Sequence 6822, Application US/10719993  
; Publication No. US20040265849A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001496  
; CURRENT APPLICATION NUMBER: US/10/719,993  
; CURRENT FILING DATE: 2003-11-24  
; NUMBER OF SEQ ID NOS: 55342  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6822  
; LENGTH: 786452  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)...(786452)  
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1  
US-10-719-993-6822

Query Match 84.0%; Score 16.8; DB 8; Length 786452;  
Best Local Similarity 90.0%; Pred. No. 6.8e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGACCAAC 20  
Db 581551 GAGCCTAACAGCAACCAAC 581570

RESULT 20  
US-10-425-115-162723/c  
; Sequence 162723, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua

```
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 162723
; LENGTH: 225
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_79980C.1
US-10-425-115-162723

Query Match      82.0%; Score 16.4; DB 8; Length 225;
Best Local Similarity 94.4%; Pred. No. 7.8e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 AGCCAAACAGGACCAAA 19
Db      116 AGCAAAACAGGACCAAA 99

RESULT 21
US-09-925-065A-873201
; Sequence 873201, Application US/09925065A
; Publication No. US2005028172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108627.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 873201
; LENGTH: 545
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-873201

Query Match      82.0%; Score 16.4; DB 4; Length 545;
Best Local Similarity 94.4%; Pred. No. 8.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 AGCCAAACAGGACCAAA 19
Db      153 AGCCAATCAGGACCAAA 170

RESULT 22
US-10-424-599-81338
; Sequence 81338, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
```

```
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 81338
; LENGTH: 570
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_44463C.1
US-10-424-599-81338

Query Match      82.0%; Score 16.4; DB 7; Length 570;
Best Local Similarity 94.4%; Pred. No. 8.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 AGCCAAACAGGACCAAA 19
Db      223 AGCCAAACAGGACCAAA 240

RESULT 23
US-10-972-079-71375/c
; Sequence 71375, Application US/10972079
; Publication No. US20050153317A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: Denise, Sue K.
; APPLICANT: ROSENFELD, David
; APPLICANT: KERR, Richard
; APPLICANT: BATES, Stephen
; APPLICANT: HOLM, Tom
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEEF
; TITLE OF INVENTION: LIVESTOCK
; FILE REFERENCE: MM1110-2
; CURRENT APPLICATION NUMBER: US/10/972,079
; CURRENT FILING DATE: 2004-10-22
; PRIOR APPLICATION NUMBER: US 60/514,333
; PRIOR FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 96631
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 71375
; LENGTH: 599
; TYPE: DNA
; ORGANISM: Chicken 19866894349953_1
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(51)
; OTHER INFORMATION: n is any nucleotide
US-10-972-079-71375

Query Match      82.0%; Score 16.4; DB 9; Length 599;
Best Local Similarity 94.4%; Pred. No. 8.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 AGCCAAACAGGACCAAA 19
Db      289 ACCCAACAGGACCAAA 272

RESULT 24
US-10-972-079-42033/c
; Sequence 42033, Application US/10972079
; Publication No. US20050153317A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: Denise, Sue K.
; APPLICANT: ROSENFELD, David
; APPLICANT: KERR, Richard
; APPLICANT: BATES, Stephen
; APPLICANT: HOLM, Tom
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEEF
; TITLE OF INVENTION: LIVESTOCK
```

```
; FILE REFERENCE: MM1110-2
; CURRENT APPLICATION NUMBER: US/10/972,079
; CURRENT FILING DATE: 2004-10-22
; PRIOR APPLICATION NUMBER: US 60/514,333
; PRIOR FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 96631
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 42033
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Chicken 19866894278533_1
US-10-972-079-42033
```

```
Query Match      82.0%; Score 16.4; DB 9; Length 600;
Best Local Similarity 85.0%; Pred. No. 8.1e+02;
Matches 17; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      1 GAGCCAAACAGGAACCAAC 20
      ||||| :|||||
Db      440 GAGCCAAACRGAACCAAC 421
```

```
RESULT 25
US-10-972-079-42034/c
; Sequence 42034, Application US/10972079
; Publication No. US20050153317A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: ROSENFELD, David
; APPLICANT: KERR, Richard
; APPLICANT: BATES, Stephen
; APPLICANT: HOLM, Tom
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEE
; FILE REFERENCE: MM1110-2
; CURRENT APPLICATION NUMBER: US/10/972,079
; CURRENT FILING DATE: 2004-10-22
; PRIOR APPLICATION NUMBER: US 60/514,333
; PRIOR FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 96631
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 42034
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Chicken 19866894278533_2
US-10-972-079-42034
```

```
Query Match      82.0%; Score 16.4; DB 9; Length 600;
Best Local Similarity 85.0%; Pred. No. 8.1e+02;
Matches 17; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      1 GAGCCAAACAGGAACCAAC 20
      ||||| :|||||
Db      439 GAGCCAAACRGAACCAAC 420
```

```
RESULT 26
US-10-972-079-42035/c
; Sequence 42035, Application US/10972079
; Publication No. US20050153317A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: ROSENFELD, David
; APPLICANT: KERR, Richard
; APPLICANT: BATES, Stephen
; APPLICANT: HOLM, Tom
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEE
; FILE REFERENCE: MM1110-2
; CURRENT APPLICATION NUMBER: US/10/972,079
; CURRENT FILING DATE: 2004-10-22
```

```
; PRIOR APPLICATION NUMBER: US 60/514,333
; PRIOR FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 96631
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 42035
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Chicken 19866894278533_3
US-10-972-079-42035
```

```
Query Match      82.0%; Score 16.4; DB 9; Length 600;
Best Local Similarity 85.0%; Pred. No. 8.1e+02;
Matches 17; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      1 GAGCCAAACAGGAACCAAC 20
      ||||| :|||||
Db      310 GAGCCAAACRGAACCAAC 291
```

```
RESULT 27
US-10-972-079-42036/c
; Sequence 42036, Application US/10972079
; Publication No. US20050153317A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: ROSENFELD, David
; APPLICANT: KERR, Richard
; APPLICANT: BATES, Stephen
; APPLICANT: HOLM, Tom
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEE
; FILE REFERENCE: MM1110-2
; CURRENT APPLICATION NUMBER: US/10/972,079
; CURRENT FILING DATE: 2004-10-22
; PRIOR APPLICATION NUMBER: US 60/514,333
; PRIOR FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 96631
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 42036
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Chicken 19866894278533_4
US-10-972-079-42036
```

```
Query Match      82.0%; Score 16.4; DB 9; Length 600;
Best Local Similarity 85.0%; Pred. No. 8.1e+02;
Matches 17; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      1 GAGCCAAACAGGAACCAAC 20
      ||||| :|||||
Db      308 GAGCCAAACRGAACCAAC 289
```

```
RESULT 28
US-10-972-079-42037/c
; Sequence 42037, Application US/10972079
; Publication No. US20050153317A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: ROSENFELD, David
; APPLICANT: KERR, Richard
; APPLICANT: BATES, Stephen
; APPLICANT: HOLM, Tom
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEE
; FILE REFERENCE: MM1110-2
; CURRENT APPLICATION NUMBER: US/10/972,079
; CURRENT FILING DATE: 2004-10-22
; PRIOR APPLICATION NUMBER: US 60/514,333
; PRIOR FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 96631
```

```
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 42037
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Chicken 19866894278533_5
US-10-972-079-42037

Query Match      82.0%; Score 16.4; DB 9; Length 600;
Best Local Similarity 85.0%; Pred. No. 8.1e+02;
Matches 17; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAAAAC 20
    ||||| :|||||
Db 197 GAGCCAAACRGAACCAAC 178

RESULT 29
US-10-972-079-42038/c
; Sequence 42038, Application US/10972079
; Publication No. US20050153317A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: ROSENFELD, David
; APPLICANT: KERR, Richard
; APPLICANT: BATES, Stephen
; APPLICANT: HOLM, Tom
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BREED
; TITLE OF INVENTION: LIVESTOCK
; FILE REFERENCE: MM1110-2
; CURRENT APPLICATION NUMBER: US/10/972,079
; CURRENT FILING DATE: 2004-10-22
; PRIOR APPLICATION NUMBER: US 60/514,333
; PRIOR FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 96631
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 42038
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Chicken 19866894278533_6
US-10-972-079-42038

Query Match      82.0%; Score 16.4; DB 9; Length 600;
Best Local Similarity 85.0%; Pred. No. 8.1e+02;
Matches 17; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAAAAC 20
    ||||| :|||||
Db 125 GAGCCAAACRGAACCAAC 106

RESULT 30
US-10-424-599-74758
; Sequence 74758, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 74758
; LENGTH: 745
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_3851C.1
US-10-424-599-74758

Query Match      82.0%; Score 16.4; DB 9; Length 600;
Best Local Similarity 85.0%; Pred. No. 8.1e+02;
Matches 17; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAAAAC 20
    ||||| :|||||
Db 125 GAGCCAAACRGAACCAAC 106

RESULT 31
US-10-424-599-80664
; Sequence 80664, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 80664
; LENGTH: 1195
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_43857C.1
US-10-424-599-80664

Query Match      82.0%; Score 16.4; DB 7; Length 1195;
Best Local Similarity 94.4%; Pred. No. 8.3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GCCAAACAGGAACCAAC 20
    ||||| |||||
Db 720 GCCAAACAGGAACCAAC 737

RESULT 32
US-09-925-065A-677481/c
; Sequence 677481, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108927.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 677481
; LENGTH: 1557
; TYPE: DNA
; ORGANISM: Homo sapiens
; OTHER INFORMATION: Clone ID: PAT_MRT3847_3851C.1
US-09-925-065A-677481

Query Match      82.0%; Score 16.4; DB 4; Length 1557;
Best Local Similarity 94.4%; Pred. No. 8.4e+02;
```

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAA 18  
|||||  
Db 1370 GAGCCAAACAGGAACCAA 1353

## RESULT 33

US-10-437-963-51993  
; Sequence 51993, Application US/10437963  
; Publication No. US20040123343A1

GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping

TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53221)B

CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14

NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 51993

LENGTH: 1794

TYPE: DNA

ORGANISM: Oryza sativa

FEATURE:

OTHER INFORMATION: Clone ID: PAT\_MRT4530\_54332C.1

US-10-437-963-51993

Query Match 82.0%; Score 16.4; DB 7; Length 1794;  
Best Local Similarity 94.4%; Pred. No. 8.5e+02;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GCCAAACAGGAACCAAAC 20  
|||||  
Db 1652 GCCAAACAGGAACCAAAC 1669

## RESULT 34

US-10-602-494-29  
; Sequence 29, Application US/10602494  
; Publication No. US20040265833A1

GENERAL INFORMATION:

APPLICANT: Cathy Lofton-Day

APPLICANT: Andrew Sledziewski

APPLICANT: Jeff Thomas

APPLICANT: Robert W. Day

APPLICANT: Lori Tonnes-Priddy

APPLICANT: Karen Cardon

TITLE OF INVENTION: Methods and nucleic acids for the analysis of colorectal cell

FILE REFERENCE: 47675-45

CURRENT APPLICATION NUMBER: US/10/602,494

CURRENT FILING DATE: 2003-06-23

NUMBER OF SEQ ID NOS: 385

SEQ ID NO 29

LENGTH: 2476

TYPE: DNA

ORGANISM: Homo Sapiens

US-10-602-494-29

Query Match 82.0%; Score 16.4; DB 8; Length 2476;  
Best Local Similarity 94.4%; Pred. No. 8.6e+02;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GCCAAACAGGAACCAAAC 20  
|||||  
Db 65 GCCAAACAGGAACCAAAC 82

## RESULT 35

US-10-437-963-23160  
; Sequence 23160, Application US/10437963  
; Publication No. US20040123343A1

GENERAL INFORMATION:

APPLICANT: La Rosa, Thomas J.

APPLICANT: Kovalic, David K.

APPLICANT: Zhou, Yihua

APPLICANT: Cao, Yongwei

APPLICANT: Wu, Wei

APPLICANT: Boukharov, Andrey A.

APPLICANT: Barbazuk, Brad

APPLICANT: Li, Ping

TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53221)B

CURRENT APPLICATION NUMBER: US/10/437,963

CURRENT FILING DATE: 2003-05-14

NUMBER OF SEQ ID NOS: 204966

SEQ ID NO 23160

LENGTH: 2978

TYPE: DNA

ORGANISM: Oryza sativa

FEATURE:

NAME/KEY: unsure

LOCATION: (1)..(2978)

OTHER INFORMATION: unsure at all n locations

FEATURE:

OTHER INFORMATION: Clone ID: PAT\_MRT4530\_28267C.1

US-10-437-963-23160

Query Match 82.0%; Score 16.4; DB 7; Length 2978;

Best Local Similarity 94.4%; Pred. No. 8.7e+02;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAA 18  
|||||  
Db 2170 GAGCCAAACAGGAACCAA 2187

## RESULT 36

US-10-428-487-11/c  
; Sequence 11, Application US/10428487  
; Publication No. US20040006780A1

GENERAL INFORMATION:

APPLICANT: RASTELLI, LUCA K.

APPLICANT: GERBER, HANS-PETER

TITLE OF INVENTION: VEGF-MODULATED GENES AND METHODS EMPLOYING THEM

FILE REFERENCE: 09800080-0103

CURRENT APPLICATION NUMBER: US/10/428,487

CURRENT FILING DATE: 2003-05-02

PRIOR APPLICATION NUMBER: 09/815,153

PRIOR FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,201

PRIOR FILING DATE: 2000-03-22

NUMBER OF SEQ ID NOS: 84

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 11

LENGTH: 301692

TYPE: DNA

ORGANISM: Homo sapiens

US-10-428-487-11

Query Match 82.0%; Score 16.4; DB 6; Length 301692;  
Best Local Similarity 94.4%; Pred. No. 1e+03;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GCCAAACAGGAACCAAAC 20  
|||||  
Db 10718 GCCAAACAGGAACCAAAC 10701

```
RESULT 37
US-10-367-094-195/c
; Sequence 195, Application US/10367094
; Publication No. US20040170982A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc Malandro
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
; FILE REFERENCE: 529452001500
; CURRENT APPLICATION NUMBER: US/10/367,094
; CURRENT FILING DATE: 2003-02-14
; NUMBER OF SEQ ID NOS: 203
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 195
; LENGTH: 310268
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-367-094-195

Query Match      82.0%; Score 16.4; DB 7; Length 310268;
Best Local Similarity 94.4%; Pred. No. 1e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3  GCCAACAGGACCAAC 20
      |||||
DB      11729  GCCAACAGGACCAAC 11712

RESULT 38
US-10-767-701-13180
; Sequence 13180, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 13180
; LENGTH: 707
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS26550_1
US-10-767-701-13180

Query Match      80.0%; Score 16; DB 7; Length 707;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5  CAAACAGGACCAAC 20
      |||||
DB      20  CAAACAGGACCAAC 35

RESULT 39
US-11-097-143-32660
; Sequence 32660, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
```

```
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 32660
; LENGTH: 5735
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-32660

Query Match      80.0%; Score 16; DB 10; Length 5735;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      4  CCAACAGGACCAAA 19
      |||||
DB      1936  CCAACAGGACCAAA 1951

RESULT 40
US-11-097-143-32659
; Sequence 32659, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 32659
; LENGTH: 8954
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-32659

Query Match      80.0%; Score 16; DB 10; Length 8954;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Oy 4 CCAAACAGGAACCAA 19  
| | | | | | | | | |  
Db 3413 CCAAACAGGAACCAA 3428

Search completed: November 21, 2005, 05:03:41  
Job time : 606.469 secs

This Page Blank (uspto)



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2005, 21:55:46 ; Search time 279.506 Seconds  
(without alignments)  
9.405 Million cell updates/sec

Title: US-10-627-757-22

Perfect score: 20

Sequence: 1 gagccaaacaggaaacaaac 20

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3196817 seqs, 65720914 residues

Total number of hits satisfying chosen parameters: 6393634

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_NA\_New.\*  
1: /cgn2\_6/ptodata/2/pubpna/us10\_NEW\_PUB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/us06\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/us07\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/us08\_NEW\_PUB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/us09\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCR\_NEW\_PUB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/us11\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/us11\_NEW\_PUB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/us11\_NEW\_PUB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/us60\_NEW\_PUB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	15	75.0	19	8	US-11-101-244-945379 Sequence 945379,
C 2	15	75.0	19	8	US-11-101-244-945437 Sequence 945437,
C 3	15	75.0	19	9	US-11-083-784-945379 Sequence 945379,
C 4	15	75.0	19	9	US-11-083-784-945437 Sequence 945437,
5	14.4	72.0	19	8	US-11-101-244-1069270 Sequence 1069270,
6	14.4	72.0	19	9	US-11-083-784-1069270 Sequence 1069270,
C 7	14.4	72.0	40439	1	US-10-993-509-1 Sequence 1, Appli
8	14.2	71.0	19	8	US-11-101-244-182007 Sequence 182007,
9	14.2	71.0	19	8	US-11-101-244-343265 Sequence 343265,
10	14.2	71.0	19	8	US-11-101-244-572172 Sequence 572172,
11	14.2	71.0	19	8	US-11-101-244-738064 Sequence 738064,
12	14.2	71.0	19	8	US-11-101-244-902399 Sequence 902399,
13	14.2	71.0	19	9	US-11-083-784-182007 Sequence 182007,
14	14.2	71.0	19	9	US-11-083-784-343265 Sequence 343265,
15	14.2	71.0	19	9	US-11-083-784-572172 Sequence 572172,
16	14.2	71.0	19	9	US-11-083-784-738064 Sequence 738064,
17	14.2	71.0	19	9	US-11-083-784-902399 Sequence 902399,
18	14.2	71.0	1194	1	US-10-957-569-53 Sequence 53, Appl
C 19	14.2	71.0	1547	1	US-10-131-8268-11 Sequence 11, Appl
C 20	14	70.0	19	8	US-11-101-244-535127 Sequence 535127,
21	14	70.0	19	8	US-11-101-244-639374 Sequence 639374,
C 22	14	70.0	19	8	US-11-101-244-773752 Sequence 773752,
C 23	14	70.0	19	8	US-11-101-244-773827 Sequence 773827,

C 24	14	70.0	19	9	US-11-083-784-535127 Sequence 535127,
25	14	70.0	19	9	US-11-083-784-639374 Sequence 639374,
C 26	14	70.0	19	9	US-11-083-784-773752 Sequence 773752,
C 27	14	70.0	19	9	US-11-083-784-773827 Sequence 773827,
C 28	14	70.0	1204	1	US-10-131-8268-505 Sequence 505, App
29	13.8	69.0	19	8	US-11-101-244-312781 Sequence 312781,
30	13.8	69.0	19	8	US-11-101-244-312896 Sequence 312896,
31	13.8	69.0	19	8	US-11-101-244-1425598 Sequence 1425598,
32	13.8	69.0	19	8	US-11-101-244-1459958 Sequence 1459958,
33	13.8	69.0	19	9	US-11-083-784-312781 Sequence 312781,
34	13.8	69.0	19	9	US-11-083-784-312896 Sequence 312896,
35	13.8	69.0	19	9	US-11-083-784-1425598 Sequence 1425598,
36	13.8	69.0	19	9	US-11-083-784-1459958 Sequence 1459958,
C 37	13.8	69.0	196	1	US-10-925-970-7 Sequence 7, Appl
C 38	13.8	69.0	364	1	US-10-925-970-16 Sequence 16, Appl
C 39	13.8	69.0	1761	1	US-10-925-970-5 Sequence 5, Appl
C 40	13.8	69.0	2315	1	US-10-925-970-4 Sequence 4, Appl
C 41	13.8	69.0	48763	1	US-10-663-794-3 Sequence 3, Appl
C 42	13.6	68.0	2331	1	US-10-925-970-2 Sequence 2, Appl
C 43	13.6	68.0	4008	1	US-10-925-970-1 Sequence 1, Appl
C 44	13.4	67.0	19	8	US-11-101-244-50974 Sequence 50974, A
C 45	13.4	67.0	19	8	US-11-101-244-495039 Sequence 495039,

ALIGNMENTS

RESULT 1  
US-11-101-244-945379/c  
; Sequence 945379, Application US/11101244  
; Publication No. US20050246794A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmacon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/101,244  
; CURRENT FILING DATE: 2005-04-07  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 945379  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-101-244-945379

Query Match 75.0%; Score 15; DB 8; Length 19;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 AAACAGGACCAAC 20  
Db 18 AAACAGGACCAAC 4

RESULT 2  
US-11-101-244-945437/c  
; Sequence 945437, Application US/11101244  
; Publication No. US20050246794A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmacon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William

```
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 945379
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-945379
```

```
Query Match 75.0%; Score 15; DB 8; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 6 AAACAGGAACCAAAAC 20
    |||||
Db 15 AAACAGGAACCAAAAC 1
```

## RESULT 3

```
US-11-083-784-945379/c
; Sequence 945379, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR APPLICATION NUMBER: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 945379
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-945379
```

```
Query Match 75.0%; Score 15; DB 9; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 6 AAACAGGAACCAAAAC 20
    |||||
Db 18 AAACAGGAACCAAAAC 4
```

## RESULT 4

```
US-11-083-784-945437/c
; Sequence 945437, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
```

```
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR APPLICATION NUMBER: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 945437
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-945437
```

```
Query Match 75.0%; Score 15; DB 9; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 6 AAACAGGAACCAAAAC 20
    |||||
Db 15 AAACAGGAACCAAAAC 1
```

## RESULT 5

```
US-11-101-244-1069270
; Sequence 1069270, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmoon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 1069270
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-1069270
```

```
Query Match 72.0%; Score 14.4; DB 8; Length 19;
Best Local Similarity 93.8%; Pred. No. 2.5e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 4 CCAAAACAGGAACCAAA 19
    |||||
Db 4 CCAAAACUGGAACCAAA 19
```

## RESULT 6

```
US-11-083-784-1069270
; Sequence 1069270, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmoon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
```

; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/083,784  
; CURRENT FILING DATE: 2005-03-18  
; PRIOR APPLICATION NUMBER: US/10/714,333  
; PRIOR FILING DATE: 2003-11-14  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 1069270  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-083-784-1069270

Query Match 72.0%; Score 14.4; DB 9; Length 19;  
Best Local Similarity 93.8%; Pred. No. 2.5e+02;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 CCAACAGGACCAAA 19  
||||| |||||||

Db 4 CCAACUGGACCAAA 19

## RESULT 7

US-10-993-509-1/c  
; Sequence 1, Application US/10993509  
; Publication No. US20050250121A1  
; GENERAL INFORMATION:  
; APPLICANT: Aerisens, Jeroen  
; APPLICANT: Athanasios, Maria  
; APPLICANT: Brain, Carlos  
; APPLICANT: Cohen, Nadine  
; APPLICANT: Dain, Bradley  
; APPLICANT: Denton, R. Rex  
; APPLICANT: Judson, Richard S.  
; APPLICANT: Ozdemir, Vural  
; APPLICANT: Reed, Carol R.  
; TITLE OF INVENTION: NTRK2 Genetic Markers Associated with Progression of Alzheimer's  
; FILE REFERENCE: 2300, 0060001  
; CURRENT APPLICATION NUMBER: US/10/993,509  
; CURRENT FILING DATE: 2004-11-22  
; PRIOR APPLICATION NUMBER: US 60/524,637  
; PRIOR FILING DATE: 2003-11-24  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 1  
; LENGTH: 40439  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (2402)..(2402)  
; OTHER INFORMATION: n is 'c' or 'g'  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (2722)..(2722)  
; OTHER INFORMATION: n is 'c' or 't'  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (2799)..(2799)  
; OTHER INFORMATION: n is 'g' or 'a'  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (3799)..(3898)  
; OTHER INFORMATION: n is 'a', 't', 'g', or 'c'

; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (6337)..(6436)  
; OTHER INFORMATION: n is 'a', 't', 'g', or 'c'  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (11729)..(11828)  
; OTHER INFORMATION: n is 'a', 't', 'g', or 'c'  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (14905)..(15004)  
; OTHER INFORMATION: n is 'a', 't', 'g', or 'c'  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (17523)..(17622)  
; OTHER INFORMATION: n is 'a', 't', 'g', or 'c'  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (19842)..(19941)  
; OTHER INFORMATION: n is 'a', 't', 'g', or 'c'  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (22316)..(22415)  
; OTHER INFORMATION: n is 'a', 't', 'g', or 'c'  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (24686)..(24785)  
; OTHER INFORMATION: n is 'a', 't', 'g', or 'c'  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (27030)..(27129)  
; OTHER INFORMATION: n is 'a', 't', 'g', or 'c'  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (29535)..(29634)  
; OTHER INFORMATION: n is 'a', 't', 'g', or 'c'  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (31929)..(32028)  
; OTHER INFORMATION: n is 'a', 't', 'g', or 'c'  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (34403)..(34502)  
; OTHER INFORMATION: n is 'a', 't', 'g', or 'c'  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (36929)..(37028)  
; OTHER INFORMATION: n is 'a', 't', 'g', or 'c'  
US-10-993-509-1

Query Match 72.0%; Score 14.4; DB 1; Length 40439;  
Best Local Similarity 93.8%; Pred. No. 2.3e+02;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 CAAACAGGACCAAAAC 20  
||||| |||||||

Db 8127 CCAACAGGACCAAAAC 8112

## RESULT 8

US-11-101-244-182007  
; Sequence 182007, Application US/11101244  
; Publication No. US20050246794A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmacon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/101,244

; CURRENT FILING DATE: 2005-04-07  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 182007  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-101-244-182007

Query Match 71.0%; Score 14.2; DB 8; Length 19;  
Best Local Similarity 84.2%; Pred. No. 3.1e+02;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAAA 19  
||| ||||| ||||| |||||  
Db 1 GAACCAACUGGAUCAAA 19

RESULT 9  
US-11-101-244-343265  
; Sequence 343265, Application US/11101244  
; Publication No. US20050246794A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmakon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/101,244  
; CURRENT FILING DATE: 2005-04-07  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 343265  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-101-244-343265

Query Match 71.0%; Score 14.2; DB 8; Length 19;  
Best Local Similarity 84.2%; Pred. No. 3.1e+02;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAAA 19  
||| ||||| ||||| |||||  
Db 1 GAUCCAAUCAGGAAGCAAA 19

RESULT 10  
US-11-101-244-572172  
; Sequence 572172, Application US/11101244  
; Publication No. US20050246794A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmakon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/101,244  
; CURRENT FILING DATE: 2005-04-07

; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 572172  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-101-244-572172

Query Match 71.0%; Score 14.2; DB 8; Length 19;  
Best Local Similarity 84.2%; Pred. No. 3.1e+02;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAAA 19  
||| ||||| ||||| |||||  
Db 1 GAGAGAAACAGGAACAAA 19

RESULT 11  
US-11-101-244-738064  
; Sequence 738064, Application US/11101244  
; Publication No. US20050246794A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmakon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/101,244  
; CURRENT FILING DATE: 2005-04-07  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 738064  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-101-244-738064

Query Match 71.0%; Score 14.2; DB 8; Length 19;  
Best Local Similarity 84.2%; Pred. No. 3.1e+02;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAAA 19  
||| ||||| ||||| |||||  
Db 1 GAAACAACAGGAAGCAAA 19

RESULT 12  
US-11-101-244-902399  
; Sequence 902399, Application US/11101244  
; Publication No. US20050246794A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmakon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/101,244  
; CURRENT FILING DATE: 2005-04-07  
; PRIOR APPLICATION NUMBER: 60/502,050

; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 902399  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-101-244-902399

Query Match 71.0%; Score 14.2; DB 8; Length 19;  
Best Local Similarity 84.2%; Pred. No. 3.1e+02;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAAA 19  
||| ||||| ||||| |||||  
Db 1 GAGGGAACAGAAACCAAA 19

RESULT 13  
US-11-083-784-182007  
; Sequence 182007, Application US/11083784  
; Publication No. US20050245475A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmacon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/083,784  
; CURRENT FILING DATE: 2005-03-18  
; PRIOR APPLICATION NUMBER: US/10/714,333  
; PRIOR FILING DATE: 2003-11-14  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 182007  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-083-784-182007

Query Match 71.0%; Score 14.2; DB 9; Length 19;  
Best Local Similarity 84.2%; Pred. No. 3.1e+02;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAAA 19  
||| ||||| ||||| |||||  
Db 1 GAACCAACUGGAACCAAA 19

RESULT 14  
US-11-083-784-343265  
; Sequence 343265, Application US/11083784  
; Publication No. US20050245475A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmacon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/083,784  
; CURRENT FILING DATE: 2005-03-18

; PRIOR APPLICATION NUMBER: US/10/714,333  
; PRIOR FILING DATE: 2003-11-14  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 343265  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-083-784-343265

Query Match 71.0%; Score 14.2; DB 9; Length 19;  
Best Local Similarity 84.2%; Pred. No. 3.1e+02;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAAA 19  
||| ||||| ||||| |||||  
Db 1 GAUCCAAUCAGGAACCAAA 19

RESULT 15  
US-11-083-784-572172  
; Sequence 572172, Application US/11083784  
; Publication No. US20050245475A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmacon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/083,784  
; CURRENT FILING DATE: 2005-03-18  
; PRIOR APPLICATION NUMBER: US/10/714,333  
; PRIOR FILING DATE: 2003-11-14  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 572172  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-083-784-572172

Query Match 71.0%; Score 14.2; DB 9; Length 19;  
Best Local Similarity 84.2%; Pred. No. 3.1e+02;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAAA 19  
||| ||||| ||||| |||||  
Db 1 GAGAGAAACAGGAACCAAA 19

RESULT 16  
US-11-083-784-738064  
; Sequence 738064, Application US/11083784  
; Publication No. US20050245475A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmacon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA

```
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR APPLICATION NUMBER: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 738064
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-738064

Query Match          71.0%; Score 14.2; DB 9; Length 19;
Best Local Similarity 84.2%; Pred. No. 3.1e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAA 19
Db 1 GAAACAAACAGGAAGCAA 19

RESULT 17
US-11-083-784-902399
; Sequence 902399, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scarings, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR APPLICATION NUMBER: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 902399
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-902399

Query Match          71.0%; Score 14.2; DB 9; Length 19;
Best Local Similarity 84.2%; Pred. No. 3.1e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAGCCAAACAGGAACCAA 19
Db 1 GAGGGAACAGAAACCAA 19

RESULT 18
US-10-957-569-53/c
; Sequence 53, Application US/10957569
; Publication No. US20050246785A1
; GENERAL INFORMATION:
; APPLICANT: COOK, Zhihong et al.
; TITLE OF INVENTION: PROMOTER, PROMOTER CONTROL ELEMENTS, AND COMBINATIONS, AND USES THEREOF
; FILE REFERENCE: 2750-1577PUS3
```

```
; CURRENT APPLICATION NUMBER: US/10/957,569
; CURRENT FILING DATE: 2004-09-30
; PRIOR APPLICATION NUMBER: US 10/950,321
; PRIOR FILING DATE: 2004-09-23
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 53
; LENGTH: 1194
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-957-569-53

Query Match          71.0%; Score 14.2; DB 1; Length 1194;
Best Local Similarity 84.2%; Pred. No. 3e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AGCCAAACAGGAACCAAC 20
Db 1082 AGCCAAAGAGAAACAAAC 1064

RESULT 19
US-10-131-826A-11/c
; Sequence 11, Application US/10131826A
; Publication No. US20050245730A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330RIC128
; CURRENT APPLICATION NUMBER: US/10/131,826A
; CURRENT FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 11
; LENGTH: 1547
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
US-10-131-826A-11

Query Match      71.0%; Score 14.2; DB 1; Length 1547;
Best Local Similarity 84.2%; Pred. No. 3e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY  2 AGCCAAACAGGAGCAAC 20
    ||||| ||||| ||||| ||
Db   670 AGCCCAACAGAACACAC 652

RESULT 20
US-11-101-244-535127/c
; Sequence 535127, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 535127
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-535127

Query Match      70.0%; Score 14; DB 8; Length 19;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  5 CAACACAGGAGCAAC 18
    ||||| ||||| ||||| ||
Db   18 CAACACAGGAGCAAC 5

RESULT 21
US-11-101-244-639374
; Sequence 639374, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 639374
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-639374

Query Match      70.0%; Score 14; DB 8; Length 19;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  7 AACAGGAGCAAC 20
    ||||| ||||| ||||| ||
Db   17 AACAGGAGCAAC 4

RESULT 22
US-11-101-244-773752/c
; Sequence 773752, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 773752
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-773752

Query Match      70.0%; Score 14; DB 8; Length 19;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  6 AACAGGAGCAAC 19
    ||||| ||||| ||||| ||
Db   1 AACAGGAGCAAC 14

RESULT 23
US-11-101-244-773827/c
; Sequence 773827, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 773827
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-773827
```

```
Query Match          70.0%; Score 14; DB 8; Length 19;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 AACAGGAACCAAC 20
    |||||
Db 17 AACAGGAACCAAC 4

RESULT 24
US-11-083-784-535127/c
; Sequence 535127, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; PRIOR FILING DATE: 2005-03-18
; PRIOR APPLICATION NUMBER: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 535127
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-535127

Query Match          70.0%; Score 14; DB 9; Length 19;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 CAAACAGGAACCAA 18
    |||||
Db 18 CAAACAGGAACCAA 5

RESULT 25
US-11-083-784-639374
; Sequence 639374, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR APPLICATION NUMBER: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 639374
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-639374

Query Match          70.0%; Score 14; DB 9; Length 19;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 AACAGGAACCAAC 20
    |||||
Db 17 AACAGGAACCAAC 4

RESULT 26
US-11-083-784-773752/c
; Sequence 773752, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR APPLICATION NUMBER: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 773752
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-773752

Query Match          70.0%; Score 14; DB 9; Length 19;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 AACAGGAACCAAC 20
    |||||
Db 17 AACAGGAACCAAC 4

RESULT 27
US-11-083-784-773827/c
; Sequence 773827, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR APPLICATION NUMBER: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 773827
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-773827
```



; SOFTWARE: Proprietary  
; SEQ ID NO 773827  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-083-784-773827

Query Match 70.0%; Score 14; DB 9; Length 19;  
Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 AACAGGAACCAAC 20  
Db 17 AACAGGAACCAAC 4

## RESULT 28

US-10-131-826A-505/c  
; Sequence 505, Application US/10131826A  
; Publication No. US20050245730A1

; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tumas, Daniel

; APPLICANT: Watanabe, Colin K

; APPLICANT: Wood, William

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; TITLE OF INVENTION: ACIDS ENCODING THE SAME

; FILE REFERENCE: P3330R1C128

; CURRENT APPLICATION NUMBER: US/10/131,826A

; CURRENT FILING DATE: 2002-04-24

; PRIOR APPLICATION NUMBER: 60/049911

; PRIOR FILING DATE: 1997-06-18

; PRIOR APPLICATION NUMBER: 60/056974

; PRIOR FILING DATE: 1997-08-26

; PRIOR APPLICATION NUMBER: 60/059113

; PRIOR FILING DATE: 1997-09-17

; PRIOR APPLICATION NUMBER: 60/059115

; PRIOR FILING DATE: 1997-09-17

; PRIOR APPLICATION NUMBER: 60/059117

; PRIOR FILING DATE: 1997-09-17

; PRIOR APPLICATION NUMBER: 60/059122

; PRIOR FILING DATE: 1997-09-17

; PRIOR APPLICATION NUMBER: 60/059184

; PRIOR FILING DATE: 1997-09-17

; PRIOR APPLICATION NUMBER: 60/059263

; PRIOR FILING DATE: 1997-09-18

; PRIOR APPLICATION NUMBER: 60/059352

; PRIOR FILING DATE: 1997-09-19

; PRIOR APPLICATION NUMBER: 60/059588

; PRIOR FILING DATE: 1997-09-19

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 550

; SEQ ID NO 505

; LENGTH: 1204

; TYPE: DNA

; ORGANISM: Homo Sapien

US-10-131-826A-505

Query Match 70.0%; Score 14; DB 1; Length 1204;  
Best Local Similarity 100.0%; Pred. No. 3.6e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 7 AACAGGAACCAAC 20  
Db 918 AACAGGAACCAAC 905

## RESULT 29

US-11-101-244-312781

; Sequence 312781, Application US/11101244

; Publication No. US20050246794A1

; GENERAL INFORMATION:

; APPLICANT: Dharmacon, Inc.

; APPLICANT: Khvorova, Anastasia

; APPLICANT: Reynolds, Angela

; APPLICANT: Leake, Devin

; APPLICANT: Marshall, William

; APPLICANT: Scaringe, Stephen

; TITLE OF INVENTION: Functional and Hyperfunctional siRNA

; FILE REFERENCE: 13499US

; CURRENT APPLICATION NUMBER: US/11/101,244

; CURRENT FILING DATE: 2005-04-07

; PRIOR APPLICATION NUMBER: 60/502,050

; PRIOR FILING DATE: 2003-09-10

; PRIOR APPLICATION NUMBER: 60/426,137

; PRIOR FILING DATE: 2002-11-14

; NUMBER OF SEQ ID NOS: 1591911

; SOFTWARE: Proprietary

; SEQ ID NO 312781

; LENGTH: 19

; TYPE: RNA

; ORGANISM: Homo sapiens

US-11-101-244-312781

Query Match 69.0%; Score 13.8; DB 8; Length 19;

Best Local Similarity 88.2%; Pred. No. 4.5e+02;

Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 CCAACAGGAACCAAC 20

Db 1 CCAACAGGAACCAAC 17

## RESULT 30

US-11-101-244-312896

; Sequence 312896, Application US/11101244

; Publication No. US20050246794A1

; GENERAL INFORMATION:

; APPLICANT: Dharmacon, Inc.

; APPLICANT: Khvorova, Anastasia

; APPLICANT: Reynolds, Angela

; APPLICANT: Leake, Devin

; APPLICANT: Marshall, William

; APPLICANT: Scaringe, Stephen

; TITLE OF INVENTION: Functional and Hyperfunctional siRNA

; FILE REFERENCE: 13499US

; CURRENT APPLICATION NUMBER: US/11/101,244

; CURRENT FILING DATE: 2005-04-07

; PRIOR APPLICATION NUMBER: 60/502,050

; PRIOR FILING DATE: 2003-09-10

; PRIOR APPLICATION NUMBER: 60/426,137

; PRIOR FILING DATE: 2002-11-14

; NUMBER OF SEQ ID NOS: 1591911

; SOFTWARE: Proprietary

; SEQ ID NO 312896

; LENGTH: 19

; TYPE: RNA

; ORGANISM: Homo sapiens

US-11-101-244-312896

Query Match 69.0%; Score 13.8; DB 8; Length 19;

Best Local Similarity 88.2%; Pred. No. 4.5e+02;

Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy	4	CCAAACAGGAACCAAC	20
Db	1	CCAAACAGAAACCAAC	17

```

RESULT 31
US-11-101-244-1425598
; Sequence 1425598, Application US/1101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Reynolds, Devin
; APPLICANT: Leake, William
; APPLICANT: Marshall, William
; APPLICANT: Scarsinge, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional sirRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101.244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 1425598
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-1425598

```

```

RESULT 32
US-11-101-244-1459958
; Sequence 1459958, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 1459958
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-1459958

```

```

Qy      1 GAGCCAACACGAGGACCA 17
        || ||||| ||||
Db      1 GAACCAACACGAGGACCA 17

RESULT 33
US-11-083-784-312781
; Sequence 312781, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scarsinge, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR APPLICATION NUMBER: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 312781
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-312781

```

```

RESULT 34
US-11-083-784-312896
; Sequence 312896, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 134990S
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR APPLICATION NUMBER: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 312896
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-312896
Query Match          69.0%;   Score 13.8;   DB 9;   Length 19;

```



```
; NAME/KEY: misc feature
; LOCATION: (1)..(364)
; OTHER INFORMATION: any n is a, g, c, t, unknown, or other
US-10-925-970-16
```

```
Query Match          69.0%; Score 13.8; DB 1; Length 364;
Best Local Similarity 88.2%; Pred. No. 4.4e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 GAGCCAAACAGGAGCCA 17
|||||
Db 32 GAGCCAAACAGGAGCCA 16
```

## RESULT 39

```
US-10-925-970-5/c
; Sequence 5, Application US/10925970
; Publication No. US20050249741A1
; GENERAL INFORMATION:
; APPLICANT: KIMURA, Toru
; APPLICANT: KIKUCHI, Kaoru
; TITLE OF INVENTION: NOVEL SEMAPHORIN GENE: SEMAPHORIN W
; FILE REFERENCE: 0020-4546P
; CURRENT APPLICATION NUMBER: US/10/925,970
; CURRENT FILING DATE: 2004-08-26
; PRIOR APPLICATION NUMBER: US/09/284,180
; PRIOR FILING DATE: 1999-06-09
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1761
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(1761)
; OTHER INFORMATION: Coding region
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(1761)
; OTHER INFORMATION: strandedness: double
US-10-925-970-5
```

```
Query Match          69.0%; Score 13.8; DB 1; Length 1761;
Best Local Similarity 88.2%; Pred. No. 4.4e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 GAGCCAAACAGGAGCCA 17
|||||
Db 953 GAGCCAAACAGGAGCCA 937
```

## RESULT 40

```
US-10-925-970-4/c
; Sequence 4, Application US/10925970
; Publication No. US20050249741A1
; GENERAL INFORMATION:
; APPLICANT: KIMURA, Toru
; APPLICANT: KIKUCHI, Kaoru
; TITLE OF INVENTION: NOVEL SEMAPHORIN GENE: SEMAPHORIN W
; FILE REFERENCE: 0020-4546P
; CURRENT APPLICATION NUMBER: US/10/925,970
; CURRENT FILING DATE: 2004-08-26
; PRIOR APPLICATION NUMBER: US/09/284,180
; PRIOR FILING DATE: 1999-06-09
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 2315
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
```

```
; LOCATION: (1)..(1764)
; OTHER INFORMATION: Coding region
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: (1765)..(2315)
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(2315)
; OTHER INFORMATION: strandedness: double
US-10-925-970-4
```

```
Query Match          69.0%; Score 13.8; DB 1; Length 2315;
Best Local Similarity 88.2%; Pred. No. 4.4e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 GAGCCAAACAGGAGCCA 17
|||||
Db 953 GAGCCAAACAGGAGCCA 937
```

```
Search completed: November 21, 2005, 05:22:41
Job time : 279.506 secs
```

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2005, 16:48:26 ; Search time 695.494 Seconds  
(without alignments)  
1552.890 Million cell updates/sec

Title: US-10-627-757-27

Perfect score: 19

Sequence: 1 ccacgtgcataccaaattga 19

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb.ba.\*

2: gb.in.\*

3: gb.env.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pr.\*

9: gb.ro.\*

10: gb.sts.\*

11: gb.sy.\*

12: gb.un.\*

13: gb.vi.\*

14: gb.htg.\*

15: gb.pl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19	100.0	19	6	CQ771179 Sequence
2	19	100.0	1116	6	CQ771160 Sequence
3	19	100.0	1913	8	AK122746 Homo sapi
4	19	100.0	2139	8	AF283519S6
5	19	100.0	196203	8	AL353555 Human DNA
6	19	100.0	208202	14	AC013446 Homo sapi
7	17.4	91.6	131184	9	AL844513 Mouse DNA
8	17.4	91.6	153493	14	AC026215 Homo sapi
9	17.4	91.6	154904	9	AC142227 Mus muscu
10	17.4	91.6	158620	14	AC069522 Homo sapi
11	17.4	91.6	160445	14	AC022376 Homo sapi
12	17.4	91.6	160722	14	AC016963 Homo sapi
13	17.4	91.6	162422	8	AC026882 Homo sapi
14	17.4	91.6	166397	5	BX530079 Zebrafish
15	17.4	91.6	170084	14	BX321918 Danio rer
16	17.4	91.6	172111	14	AC018498 Homo sapi
17	17.4	91.6	186007	8	AC024057 Homo sapi
18	17.4	91.6	209214	9	AL844164 Mouse DNA

19	17.4	91.6	212916	14	AC096939 Rattus no
20	17.4	91.6	228412	14	AC135524 Rattus no
c 21	17.4	91.6	237447	14	AC134630 Rattus no
c 22	17.4	91.6	237900	14	AC132525 Rattus no
c 23	17.4	91.6	244647	14	AC126628 Rattus no
c 24	17.4	91.6	251288	14	AC097593 Rattus no
c 25	17.4	91.6	276127	14	AC116265 Rattus no
c 26	17.4	91.6	278063	14	AC123403 Rattus no
c 27	17.4	91.6	314016	14	AC146012 Pan trogl
c 28	17	89.5	3242	8	AK097742 Homo sapi
c 29	17	89.5	88176	8	AL137802 Human DNA
c 30	17	89.5	110000	14	AL713858 Homo sapi
c 31	17	89.5	144605	14	AC015618 Homo sapi
c 32	17	89.5	170869	14	AL691480 Mus muscu
c 33	17	89.5	203055	14	AL355800 Homo sapi
c 34	17	89.5	275252	14	AC161653 Bos tauru
c 35	16.4	86.3	201	10	BV201803 sqm20743
c 36	16.4	86.3	359	6	CS070928 Sequence
c 37	16.4	86.3	465	6	AX407370 Sequence
c 38	16.4	86.3	711	6	CQ612015 Sequence
c 39	16.4	86.3	1404	6	AR378249 Sequence
c 40	16.4	86.3	1557	8	BC022319 Homo sapi
c 41	16.4	86.3	1879	8	BC017241 Homo sapi
c 42	16.4	86.3	1973	6	CS071136 Sequence
c 43	16.4	86.3	2546	8	BC008351 Homo sapi
c 44	16.4	86.3	2555	8	BC001738 Homo sapi
c 45	16.4	86.3	2711	6	CQ612014 Sequence

## ALIGNMENTS

RESULT 1	CQ771179	CQ771179	19 bp	DNA	linear	PAT 04-MAR-2004
LOCUS	Sequence 27 from Patent EP1388590.					
DEFINITION	CQ771179					
ACCESSION	CQ771179.1	GI:45125312				
VERSION						
KEYWORDS	synthetic construct					
SOURCE	synthetic construct					
ORGANISM	other sequences; artificial sequences.					
REFERENCE	1					
AUTHORS	Kouchi,Y., Masago,A. and Takahata,T.					
TITLE	Gene assay method for predicting glaucoma onset risk					
JOURNAL	Patent: EP 1388590-A 27 11-FEB-2004;					
FEATURES	Synex Corporation (JP)					
source	Location/Qualifiers					
	1..19					
	/organism="synthetic construct"					
	/mol_type="unassigned DNA"					
	/db_xref="taxon:33630"					

## ORIGIN

Query Match	100.0%;	Score 19;	DB 6;	Length 19;
Best Local Similarity	100.0%;	Pred. No. 15;		
Matches	19;	Conservative	0;	Mismatches
			0;	Indels
				Gaps
				0;
QY	1	CCACGTGCATCCAAATTGA	19	
DB	1	CCACGTGCATCCAAATTGA	19	
RESULT 2	CQ771160	CQ771160	1116 bp	DNA
LOCUS	Sequence 8 from Patent EP1388590.			
DEFINITION	CQ771160			
ACCESSION	CQ771160.1	GI:45125293		
VERSION				
KEYWORDS	Homo sapiens (human)			
SOURCE	Homo sapiens			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			

```

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
AUTHORS Kouchi,Y., Masago,A. and Takahata,T.
TITLE Gene assay method for predicting glaucoma onset risk
JOURNAL Patent: EP 1388590-A 8 11-FEB-2004;
SYNOPSIS Sysmex Corporation (JP)
FEATURES
source Location/Qualifiers
1. .1116
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 100.0%; Score 19; DB 6; Length 1116;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCAGTGCATCCAAATTGA 19
|||||
Db 251 CCAGTGCATCCAAATTGA 269
RESULT 3
AK122746
LOCUS Homo sapiens cDNA FLJ16271 fis, clone NOVAP002039, highly similar
to Homo sapiens FIP2 alternatively translated mRNA.
ACCESSION AK122746.1 GI:34527941
VERSION AK122746.1
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1
AUTHORS Ota,T., Suzuki,Y., Nishikawa,T., Otsuki,T., Sugiyama,T., Irie,R.,
Wakamatsu,A., Hayashi,K., Sato,H., Nagai,K., Kimura,K., Makita,H.,
Sekine,M., Oobayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
Ishii,S., Yamamoto,J., Saito,K., Kawai,Y., Isono,Y., Nakamura,Y.,
Nagahari,K., Murakami,K., Yasuda,T., Iwayanagi,T., Wagatsuma,M.,
Shiratori,A., Sudo,H., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
Sugawara,M., Takahashi,M., Kanda,K., Yokoi,T., Furuya,T.,
Kikkawa,E., Omura,Y., Abe,K., Kamihara,K., Katsuta,N., Sato,K.,
Tanikawa,M., Yamazaki,M., Ninomiya,K., Ishibashi,T., Yamashita,H.,
Murakawa,K., Fujimori,K., Tanai,H., Kimata,M., Watanabe,M.,
Hiraoka,S., Chiba,Y., Ishida,S., Ono,Y., Takiguchi,S., Watanabe,S.,
Yosida,M., Hotta,T., Kusano,J., Kanehori,K., Takahashi-Fujii,A.,
Hara,H., Tanase,T., Nomura,Y., Togiya,S., Komai,F., Hara,R.,
Takeuchi,K., Arita,M., Imose,N., Musashino,K., Yuki,H., Oshima,A.,
Sasaki,N., Aotsuka,S., Yoshikawa,Y., Matsunawa,H., Ichihara,T.,
Shiohata,N., Sano,S., Moriya,S., Momiyama,H., Satoh,N., Takami,S.,
Terashima,Y., Suzuki,O., Nakagawa,S., Senoh,A., Mizoguchi,H.,
Goto,Y., Shimizu,F., Wakebe,H., Hishigaki,H., Watanabe,T.,
Sugiyama,A., Takenoto,M., Kawakami,B., Yamazaki,M., Watanabe,K.,
Kumagai,A., Itakura,S., Fukuzumi,Y., Fujimori,Y., Komiyama,M.,
Tashiro,H., Tanigami,A., Fujiwara,T., Ono,T., Yamada,K., Fujii,Y.,
Ozaki,K., Hirao,M., Ohmori,Y., Kawabata,A., Hikiji,T., Kobatake,N.,
Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Kawakami,T.,
Noguchi,S., Itoh,T., Shigetani,K., Senba,T., Matsumura,K.,
Nakajima,Y., Mizuno,T., Morinaga,M., Sasaki,M., Togashi,T.,
Oyama,M., Hata,H., Wakabe,M., Komatsu,T., Mizushima-Sugano,J.,
Satoh,T., Shirai,Y., Takahashi,Y., Nakagawa,K., Okumura,K.,
Nagase,T., Nomura,N., Kikuchi,H., Masuho,Y., Yamashita,R.,
Nakai,K., Yada,T., Nakamura,Y., Ohara,O., Isogai,T. and Sugano,S.
Complete sequencing and characterization of 21,243 full-length
human cDNAs
Nat. Genet. 36 (1), 40-45 (2004)
14702039
REFERENCE
2
AUTHORS Suzuki,O., Sasaki,N., Aotsuka,S., Shoji,T., Ichihara,T.,
Shiohata,N., Matsumoto,K., Hirano,M., Sano,S., Nomura,R.,

```

```

Yoshikawa,Y., Matsumura,Y., Moriya,S., Chiba,E., Momiyama,H.,
Onogawa,S., Kaeriyama,S., Satoh,N., Matsumura,H., Takahashi,E.,
Kataoka,R., Kuga,N., Kuroda,A., Satoh,I., Kamata,K., Takami,S.,
Terashima,Y., Watanabe,M., Sugiyama,T., Irie,R., Otsuki,T.,
Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,
Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,
Watsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
NEDO human cDNA sequencing project
3 (bases 1 to 1913)
Isogai,T. and Yamamoto,J.
Direct Submission
Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.
Location/Qualifiers
1. .1913
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="NOVAR2002039"
/tissue_type="ovary"
/clone_lib="NOVAR2"
/dev_stage="adult"
/note="cloning vector: pME18SFL3"
ORIGIN
Query Match 100.0%; Score 19; DB 8; Length 1913;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCAGTGCATCCAAATTGA 19
|||||
Db 380 CCAGTGCATCCAAATTGA 398
RESULT 4
AF283519S6
LOCUS Homo sapiens FIP2 gene, exons 9 and 10.
DEFINITION AF283524
ACCESSION AF283524.1 GI:9837252
VERSION AF283524.1
KEYWORDS 6 of 9
SEGMENT
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 (bases 1 to 2139)
Li,D. and Roberts,R.
Human FIP-2: Genomic structure and mutational analysis in ARVD
patients
Unpublished
JOURNAL 2 (bases 1 to 2139)
REFERENCE 2 (bases 1 to 2139)
Li,D. and Roberts,R.
Direct Submission
JOURNAL Submitted (30-JUN-2000) Medicine, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
Location/Qualifiers
1. .2139
/organism="Homo sapiens"

```

		/mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="10" 344..446 /number=9 1956..2071 /number=10		100.0%; Score 19; DB 8; Length 2139; Best Local Similarity 100.0%; Pred. No. 17; Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.	
exon						FEATURES	
exon						Source	
ORIGIN						Location/Qualifiers	
Query Match						1..196203	
Best Local Similarity						/organism="Homo sapiens"	
Matches						/mol_type="genomic DNA"	
						/db_xref="taxon:9606"	
						/chromosome="10"	
						/clone="RP11-730A19"	
						/clone_lib="RPCI-11.3"	
Qy		1 CCAGTGCATCCAAATTGA 19		misc_feature		/note="Clone left end: RP11-730A19"	
Db		1706 CCCAGTGCATCCAAATTGA 1724		gene		/join(complement(123034..123180), complement(122612..122837),complement(122053..122130), complement(97396..97628),complement(73202..73469), complement(21866..22040), complement(AL353586.14:43768..45820))	
RESULT 5						/gene="RP11-347122.1"	
ALJ55355						/locus_tag="RP11-347122.1-001"	
LOCUS						/join(complement(123034..123180), complement(122612..122837),complement(122053..122130), complement(97396..97628),complement(73202..73469), complement(21866..22040), complement(AL353586.14:43768..45820))	
DEFINITION						/gene="RP11-347122.1"	
						/locus_tag="RP11-347122.1-001"	
						/product="novel protein"	
						/note="match: cDNAs: AK095792.1"	
						/join(complement(24725..25225),complement(21866..22040), complement(AL353586.14:43768..45820))	
						/gene="RP11-347122.1"	
						/locus_tag="RP11-347122.1-002"	
						/join(complement(24725..25225),complement(21866..22040), complement(AL353586.14:43768..45820))	
						/gene="RP11-347122.1"	
						/locus_tag="RP11-347122.1-002"	
						/product="novel protein"	
						/note="match: ESTs: AL514062.1 AL532821.1 AL552108.1 AL571318.1 AL571347.1 AW965893.1 BE299213.1 BF307636.1 BF339232.1 BF582825.1 BM503747.1 BM677538.1 BM983097.1 BQ044480.1 BQ331824.1 BQ447486.1 BQ720316.1 BQ722244.1 BQ723478.1 BQ894764.1 BQ897889.1 BU726400.1 BU901542.1 match: cDNAs: AB050523.1 AK009833.1 AL136562.1"	
						/join(complement(24725..25098),complement(21866..22040), complement(AL353586.14:45557..45820))	
						/gene="RP11-347122.1"	
						/locus_tag="RP11-347122.1-002"	
						/standard_name="OTTHUMP0000019122"	
						/note="match: proteins: Q9BQ14 Q9D6Y1"	
						/codon_start=1	
						/product="novel protein"	
						/protein_id="CA116546.1"	
						/db_xref="GI:55958689"	
						/translation="MLRQLLAALCLAGPPAPARACOLPSEWRPLSEGCRAELAEITV YARVIALHPEAPGLYNYHPVWYHAGQGLFYSAEVMCLDQAGMLVPPAGSRNLIT GLVFSCHSHTVVDYVYFFFLRMNDENYLLPHGVNFODAI FPDQENRMFSSSLFOF SNCSGQQLATFSSDWEIQEDSRMLCQSVQKALFEEDHVKLQKVATLEKRNQLR BRVKKVKSIRQARKKGRHLELANQKSEKLAAGALPHINARGPVRPPYLRG"	
						/join(complement(21866..22039), complement(AL353586.14:45557..45820))	
						/gene="RP11-347122.1"	
						/locus_tag="RP11-347122.1-001"	
						/standard_name="OTTHUMP0000019121"	
						/codon_start=1	
						/product="novel protein"	
						/protein_id="CA116545.1"	
						/db_xref="GI:55958688"	
						/db_xref="UniProt/TrEMBL:Q5VVV9"	
						/translation="MDENYLLPHGVNFODAI FPDQENRMFSSSLFOFNSCSGQQL ATFSDWEIQEDSRMLCQSVQKALFEEDHVKLQKVATLEKRNQRNRVKKVGRS LRQARKKGRHLELANQKSEKLAAGALPHINARGPVRPPYLRG"	
						59854..59914	
						/note="Sequence from overlapping clone RP11-141H1"	
						misc_feature	
				</			

```

(AL596124) . Assembly confirmed by restriction digest.
complement(81430..82204)
/locus_tag="RP11-730A19.2-001"
/pseudo
complement(81430..82204)
/locus_tag="RP11-730A19.2-001"
/notes="match: proteins: AAN05603 Q22608 O44248 O65353
O76190 P09895 P15125 P19949 P22451 P46777 P47962
P49405 Q8MTN3 Q90YW5 Q90YW6 Q95276 Q9AVH1 Q9H3F4
Q9WSR8 Q9Y0H6"
/pseudo
/codon_start=1
/product="ribosomal protein L5 (RPL5) pseudogene"
98305..98335
/notes="Sequence from overlapping clone RP11-141H1
(AL596124) . Assembly confirmed by restriction digest."
121303
/notes="Clone right end: RP11-141H1"
join(122977..123486,131666..131817,132640..132816,
133802..134004,135981..136163,139795..139868,
142416..142568,145913..146015,147523..147638,
148946..149095,149474..149567,151273..151431,
155595..155725,157030..157109,160273..161819)
/gene="OPTN"
/locus_tag="RP11-730A19.1-006"
join(122977..123486,131666..131817,132640..132816,
133802..134004,135981..136163,139795..139868,
142416..142568,145913..146015,147523..147638,
148946..149095,149474..149567,151273..151431,
155595..155725,157030..157109,160273..161819)
/gene="OPTN"
/locus_tag="RP11-730A19.1-006"
/product="optineurin"
/notes="match: ESTs: AA196381.1 AA442279.1 A1675996.1
A1760653.1 A1937171.1 AL134550.1 AL562664.1 AL565831.1
AL583163.1 AL708106.1 AW156882.1 AW162062.1
AW162138.1 AW297795.1 AW630059.1 BE875862.1 BE884512.1
BF195338.1 BG026508.1 BG031732.1 BGI19942.1 BG222549.1
BG427366.1 BG530095.1 BG678027.1 BHS60200.1 BI770911.1
BM709706.1 BM809924.1 BM952478.1 BQ719953.1 BU623376.1
BU628287.1 BU733648.1 CA416446.1 N23490.1
match: cDNAs: AK055403.1"
join(123110..123486,131666..131817,132640..132816,
133802..134004,135981..136163,139795..139868,
142434..142568,145913..146015,147523..147638,
148946..149095,149474..149567,151273..151431,
155595..155725,157030..157109,160273..161819)
/gene="OPTN"
/locus_tag="RP11-730A19.1-009"
join(123110..123486,131666..131817,132640..132816,
133802..134004,135981..136163,139795..139868,
142434..142568,145913..146015,147523..147638,
148946..149095,149474..149567,151273..151431,
155595..155725,157030..157109,160273..161819)
/gene="OPTN"
/locus_tag="RP11-730A19.1-009"
/product="optineurin"
/notes="match: cDNAs: AB063036.1"
join(123687..123830,132640..132816,133001..133190)
/gene="OPTN"
/locus_tag="RP11-730A19.1-010"
join(123687..123830,132640..132816,133001..133190)
/gene="OPTN"
/locus_tag="RP11-730A19.1-010"
/product="optineurin"
/notes="match: ESTs: AJ182845.1 AA187695.1"
join(123691..123830,132640..132816,135981..136163,
139795..139868,142416..142508)
/gene="OPTN"
/locus_tag="RP11-730A19.1-002"
join(123691..123830,132640..132816,135981..136163,
139795..139868,142416..142508)
/gene="OPTN"

/locus_tag="RP11-730A19.1-002"
Query Match 100.0%; Score 19; DB 8; Length 196203;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATTGA 19
|||||
Db 147273 CCCAGTGCATCCAAATTGA 147291

RESULT 6
AC013446/c
LOCUS AC013446 208202 bp DNA linear HTG 07-JUL-2000
DEFINITION Homo sapiens chromosome 10 clone RP11-513P21, WORKING DRAFT
SEQUENCE, 23 unordered pieces.
AC013446
AC013446.3 GI:7923997
HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 (bases 1 to 208202)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
JOURNAL
REFERENCE
2 (bases 1 to 208202)
Waterston,R.H.
Direct Submission
AUTHORS
TITLE
JOURNAL
COMMENT
On May 18, 2000 this sequence version replaced gi.6850545.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0513P21
----- Summary Statistics -----
Sequencing vector: M13; 84%
Sequencing vector: plasmid; 16%
Chemistry: Dye-primer ET; 84% of reads
Chemistry: Dye-terminator Big Dye; 18% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 196339 bases at least Q40
Consensus quality: 199886 bases at least Q30
Consensus quality: 202038 bases at least Q20
Insert size: 215000; agarose-fp
Insert size: 206002; sum-of-contigs
Quality coverage: 4.38 in Q20 bases; agarose-fp
Quality coverage: 4.26 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 23 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1699: contig of 1699 bp in length
* 1700 1799: gap of unknown length
* 1800 3240: contig of 1441 bp in length
* 3241 3340: gap of unknown length
* 3341 6215: contig of 2875 bp in length
* 6216 6316: gap of unknown length
* 6316 9055: contig of 2740 bp in length
* 9056 9155: gap of unknown length
* 9156 11844: contig of 2689 bp in length
*

```



```

* 11845 11944: gap of unknown length
* 11945 15591: contig of 3647 bp in length
* 15592 15691: gap of unknown length
* 15692 18257: contig of 2566 bp in length
* 18258 18357: gap of unknown length
* 18358 23438: contig of 5081 bp in length
* 23439 23538: gap of unknown length
* 23539 27554: contig of 4015 bp in length
* 27554 27653: gap of unknown length
* 27654 32489: contig of 4835 bp in length
* 32490 32588: gap of unknown length
* 32589 36405: contig of 3817 bp in length
* 36406 36505: gap of unknown length
* 36506 40593: contig of 4088 bp in length
* 40594 40694: gap of unknown length
* 40694 47782: contig of 7089 bp in length
* 47783 47882: gap of unknown length
* 47883 55184: contig of 7302 bp in length
* 55185 55284: gap of unknown length
* 55285 63329: contig of 8044 bp in length
* 63330 63428: gap of unknown length
* 63429 74101: contig of 10673 bp in length
* 74102 74201: gap of unknown length
* 74202 85050: contig of 10849 bp in length
* 85051 85150: gap of unknown length
* 85151 100569: contig of 15419 bp in length
* 100570 100669: gap of unknown length
* 100670 116237: contig of 15567 bp in length
* 116238 116336: gap of unknown length
* 116337 131648: contig of 15312 bp in length
* 131649 131748: gap of unknown length
* 131749 148827: contig of 17078 bp in length
* 148828 148927: gap of unknown length
* 148928 175149: contig of 26223 bp in length
* 175150 175249: gap of unknown length
* 175250 208202: contig of 32953 bp in length.
FEATURES             Location/Qualifiers
source               1..208202
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
                    /chromosomes="10"
                    /clone="RP11-513P21"
misc_feature         1..1699
                    /note="assembly_name:Contig10"
gap                 1700..1799
                    /estimated_length=unknown
misc_feature         1800..3240
                    /note="assembly_name:Contig11"
gap                 3241..3340
                    /estimated_length=unknown
misc_feature         3341..6215
                    /note="assembly_name:Contig12"
gap                 6216..6315
                    /estimated_length=unknown
misc_feature         6316..9055
                    /note="assembly_name:Contig13"
gap                 9056..9155
                    /estimated_length=unknown
misc_feature         9156..11844
                    /note="assembly_name:Contig14"
gap                 11845..11944
                    /estimated_length=unknown
misc_feature         11945..15591
                    /note="assembly_name:Contig15"
gap                 15592..15691
                    /estimated_length=unknown
misc_feature         15692..18257
                    /note="assembly_name:Contig16"
gap                 18258..18357
                    /estimated_length=unknown
misc_feature         18358..23438
                    /note="assembly_name:Contig17"

```

```

gap                 23439..23538
                    /estimated_length=unknown
misc_feature         23539..27553
                    /note="assembly_name:Contig18"
gap                 27554..27653
                    /estimated_length=unknown
misc_feature         27654..32488
                    /note="assembly_name:Contig19"
gap                 32489..32588
                    /estimated_length=unknown
misc_feature         32589..36405
                    /note="assembly_name:Contig20"
gap                 36406..36505
                    /estimated_length=unknown
misc_feature         36506..40593
                    /note="assembly_name:Contig21"
gap                 40594..40693
                    /note="assembly_name:Contig22"
misc_feature         40694..47782
                    /estimated_length=unknown
gap                 47783..47882
                    /estimated_length=unknown
misc_feature         47883..55184
                    /note="assembly_name:Contig23"
gap                 55185..55284
                    /estimated_length=unknown
misc_feature         55285..63328
                    /note="assembly_name:Contig24"
gap                 63329..63428
                    /estimated_length=unknown
misc_feature         63429..74101
                    /note="assembly_name:Contig25"
gap                 74102..74201
                    /estimated_length=unknown
misc_feature         74202..85050
                    /note="assembly_name:Contig26"
gap                 85051..85150
                    /estimated_length=unknown
misc_feature         85151..100569
                    /note="assembly_name:Contig27"
gap                 100570..100669
                    /estimated_length=unknown
misc_feature         100670..116236
                    /note="assembly_name:Contig28"
gap                 116237..116336
                    /estimated_length=unknown
misc_feature         116337..131648
                    /note="assembly_name:Contig29"
gap                 131649..131748
                    /estimated_length=unknown
misc_feature         131749..148826
                    /note="assembly_name:Contig30"
gap                 148827..148926
                    /estimated_length=unknown
misc_feature         148927..175149
                    /note="assembly_name:Contig31"
gap                 175150..175249
                    /estimated_length=unknown
misc_feature         175250..208202
                    /note="assembly_name:Contig32"

```

## ORIGIN

```

Query Match          100.0%; Score 19; DB 14; Length 208202;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 CCCAGTGCATCCCAATTGA 19
        |||||
Db      58993 CCCAGTGCATCCCAATTGA 58975

```

```

RESULT 7
AL844513      131184 bp   DNA      linear      ROD 18-NOV-2002
LOCUS
DEFINITION
  Mouse DNA sequence from clone Rp23-132A3 on chromosome 2, complete
  sequence.
ACCESSION
AL844513
VERSION
AL844513.8   GI:33438736
KEYWORDS
HTG.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Muridae; Murinae; Mus.
1 (bases 1 to 131184)
Direct Submission
Submitted (19-NOV-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Aug 5, 2003 this sequence version replaced gi:24816942.
Sequence from the Mouse Genome Sequencing Consortium whole genome
shotgun may have been used to confirm this sequence. Sequence data
from the whole genome shotgun alone has only been used where it has
a phred quality of at least 30.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep
from the RPI-23 Mouse PAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBac3.6.
FEATURES
    source
        1..131184
            /organism="Mus musculus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10090"
            /chromosome="2"
            /clone="RP23-132A3"
            /clone_lib="RPI-23"
ORIGIN
Query Match      91.6%; Score 17.4; DB 9; Length 131184;
Best Local Similarity 94.7%; Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CCCAGTGCATCCCAATTGA 19
    |||||
Db 105272 CCCAGTGCATCCCAATTAGA 105290

```

```

RESULT 8
AC026215
LOCUS
DEFINITION
  Homo sapiens chromosome 3 clone RP11-762012 map 3p, WORKING DRAFT
  SEQUENCE, 14 unordered pieces.
ACCESSION
AC026215
VERSION
AC026215.2   GI:8101277
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 153493)
Direct Submission
Submitted (21-MAR-2000) Human Genomic Center, Institute of
Genetics, Chinese Academy of Sciences, Datun Road, Beijing
100101, P.R. China
On May 29, 2000 this sequence version replaced gi:7272032.
-----Genome Center
Center:Beijing Center
Center code:Beijing
Website:http://hgsc.igtp.ac.cn
http://www.genomics.org.cn
Contact:hgsc@igtp.ac.cn
----- Project Information
Center project name:1% project
Center clone name: RP11-762012
----- Summary Statistics
Sequencing vector: pUC18; 100% of reads
Chemistry: Dye-terminator; RT 55% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 140308 bases at least Q40
Consensus quality: 144732 bases at least Q30
Consensus quality: 148467 bases at least Q20
Insert size: 145483; sum-of-contigs
Quality coverage: 4.31x in Q20 bases;sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1354: contig of 1354 bp in length
* 1355 1454: gap of unknown length
* 1455 2848: contig of 1394 bp in length
* 2849 5925: gap of unknown length
* 5926 6025: contig of 2977 bp in length
* 6025: gap of unknown length

```

\* 6026 10256: contig of 4231 bp in length  
 \* 10257 gap of unknown length  
 \* 16217: contig of 5861 bp in length  
 \* 16218 gap of unknown length  
 \* 16317: gap of unknown length  
 \* 23323: contig of 7006 bp in length  
 \* 23324 gap of unknown length  
 \* 23424 gap of unknown length  
 \* 33387: contig of 9964 bp in length  
 \* 33487: gap of unknown length  
 \* 33488 gap of unknown length  
 \* 34088 45037: contig of 11550 bp in length  
 \* 45038 45137: gap of unknown length  
 \* 45138 58530: contig of 13393 bp in length  
 \* 58531 58630: gap of unknown length  
 \* 58631 72135: contig of 13505 bp in length  
 \* 72136 72235: gap of unknown length  
 \* 72236 85658: contig of 13423 bp in length  
 \* 85659 85758: gap of unknown length  
 \* 85759 98971: contig of 13213 bp in length  
 \* 98971 99071: gap of unknown length  
 \* 98972 122935: contig of 23864 bp in length  
 \* 122936 123035: gap of unknown length  
 \* 123036 153493: contig of 30458 bp in length.

## FEATURES

## source

1. .153493  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="3"  
 /map="3p"  
 /clone="RP11-762012"  
 1. .1354  
 /note="assembly\_name:Contig5"  
 1355. .1454  
 /estimated\_length=unknown  
 1455. .2848  
 /note="assembly\_name:Contig6"  
 2849. .2948  
 /estimated\_length=unknown  
 2949. .5925  
 /note="assembly\_name:Contig7"  
 5926. .6025  
 /estimated\_length=unknown  
 6026. .10256  
 /note="assembly\_name:Contig8"  
 10257. .10356  
 /estimated\_length=unknown  
 10357. .16217  
 /note="assembly\_name:Contig9"  
 16218. .16317  
 /estimated\_length=unknown  
 16318. .23323  
 /note="assembly\_name:Contig10"  
 23324. .23423  
 /estimated\_length=unknown  
 23424. .33387  
 /note="assembly\_name:Contig11"  
 33388. .33487  
 /estimated\_length=unknown  
 33488. .45037  
 /note="assembly\_name:Contig12"  
 45038. .45137  
 /estimated\_length=unknown  
 45138. .58530  
 /note="assembly\_name:Contig13"  
 58531. .58630  
 /estimated\_length=unknown  
 58631. .72135  
 /note="assembly\_name:Contig14"  
 72136. .72235  
 /estimated\_length=unknown  
 72236. .85658  
 /note="assembly\_name:Contig15"  
 85659. .85758  
 /estimated\_length=unknown

misc\_feature 85759. .98971  
 /note="assembly\_name:Contig16"  
 gap 98972. .99071  
 /estimated\_length=unknown  
 misc\_feature 99072. .122935  
 /note="assembly\_name:Contig17"  
 gap 122936. .123035  
 /estimated\_length=unknown  
 misc\_feature 123036. .153493  
 /note="assembly\_name:Contig18"

## ORIGIN

Query Match 91.8%; Score 17.4; DB 14; Length 153493;  
 Best Local Similarity 94.7%; Pred. No. 1.6e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATTGA 19

Db 66425 CCCAGTTCATCCAAATTGA 66443

## RESULT 9

AC142227 154904 bp DNA linear ROD 01-JAN-2004  
 LOCUS Mus musculus BAC clone RP24-115G21 from chromosome 12, complete  
 DEFINITION sequence.  
 ACCESSION AC142227  
 VERSION AC142227.3 GI:38259261  
 KEYWORDS HTG.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridea; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 154904)  
 AUTHORS Tomlinson,C., Bielicki,L. and Haakenson,W.  
 TITLE The sequence of Mus musculus BAC clone RP24-115G21  
 JOURNAL Unpublished (2001)  
 REFERENCE 2 (bases 1 to 154904)  
 AUTHORS Wilson,R.  
 TITLE Sequencing of Mus musculus  
 JOURNAL Unpublished (2001)  
 REFERENCE 3 (bases 1 to 154904)  
 AUTHORS McPherson,J.D. and Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (26-MAR-2003) Genome Sequencing Center, 4444 Forest Park  
 Parkway, St. Louis, MO 63108, USA  
 REFERENCE 4 (bases 1 to 154904)  
 AUTHORS McPherson,J.D. and Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (29-MAR-2003) Genome Sequencing Center, 4444 Forest Park  
 Parkway, St. Louis, MO 63108, USA  
 REFERENCE 5 (bases 1 to 154904)  
 AUTHORS Wilson,R.K.  
 TITLE Direct Submission  
 JOURNAL Submitted (12-NOV-2003) Genome Sequencing Center, 4444 Forest Park  
 Parkway, St. Louis, MO 63108, USA  
 REFERENCE 6 (bases 1 to 154904)  
 AUTHORS Wilson,R.  
 TITLE Direct Submission  
 JOURNAL Submitted (01-JAN-2004) Department of Genetics, Washington  
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 COMMENT On Nov 12, 2003 this sequence version replaced gi:29367027.  
 ----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu>  
 Contact: [submissions@watson.wustl.edu](mailto:submissions@watson.wustl.edu)  
 ----- Summary Statistics  
 -----  
 Center project name: M\_BB0115G21  
 -----

NOTICE: This sequence may not represent the entire insert of this

clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

#### SOURCE INFORMATION:

The RPCI-24 BAC Library has been constructed by Pieter de Jong and coworkers (<http://www.chori.org>) from male C57BL/6J mouse spleen and/or brain genomic DNA. The clone and detailed information can be obtained from Pieter de Jong and coworkers at <http://www.chori.org>

#### NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone. This clone is overlapped by AC125535.

#### FEATURES

Location/Qualifiers	
source	1. .154904
	/organism="Mus musculus"
	/mol_type="genomic DNA"
	/db_xref="taxon:10090"
	/chromosome="12"
	/map="12"
	/clone="RP24-115G21"
	/clone_lib="RPCI-24"
repeat_region	3515..4184
	/rpt_family="L1"
repeat_region	4193..4287
	/rpt_family="L1"
repeat_region	5602..6002
	/rpt_family="MaLR"
repeat_region	6157..6173
	/rpt_family="B4"
repeat_region	6174..6289
	/rpt_family="Alu"
repeat_region	6290..6303
	/rpt_family="B4"
repeat_region	7413..7776
	/rpt_family="MaLR"
repeat_region	9814..9960
	/rpt_family="Alu"
repeat_region	10260..10495
	/rpt_family="MaLR"
repeat_region	10423..10540
	/rpt_family="B4"
repeat_region	10899..10944
	/rpt_family="B4"
repeat_region	12401..12437
	/rpt_family="L1"
repeat_region	13038..13357
	/rpt_family="MaLR"
repeat_region	13384..14336
	/rpt_family="MaLR"
repeat_region	14693..14776
	/rpt_family="Alu"
repeat_region	18520..18614
	/rpt_family="B4"
repeat_region	18615..18777
	/rpt_family="B4"
repeat_region	18851..19252
	/rpt_family="MaLR"
repeat_region	19260..19324
	/rpt_family="B4"
repeat_region	19332..19612
	/rpt_family="B4"
repeat_region	19915..20047
	/rpt_family="Alu"
repeat_region	20948..21035
	/rpt_family="L1"
repeat_region	21180..21325
	/rpt_family="Alu"
repeat_region	21525..21638
	/rpt_family="Alu"
repeat_region	22319..22436
	/rpt_family="Alu"
repeat_region	22700..22789
	/rpt_family="MER1_type"
repeat_region	23207..23377
	/rpt_family="B4"
repeat_region	23610..23754
	/rpt_family="B4"
repeat_region	23755..24081
	/rpt_family="MaLR"
repeat_region	24082..24104
	/rpt_family="B4"
repeat_region	24844..25012
	/rpt_family="L1"
misc_feature	24880..24881
	/note="Bacterial transposon insertion in clone excised here."
repeat_region	25671..25718
	/rpt_family="ERV1"
repeat_region	26168..26314
	/rpt_family="Alu"
repeat_region	26362..26495
	/rpt_family="B4"
repeat_region	26902..27283
	/rpt_family="MaLR"
repeat_region	27486..27938
	/rpt_family="L1"
repeat_region	28136..28416
	/rpt_family="B4"
repeat_region	28733..29072
	/rpt_family="MaLR"
repeat_region	30225..30393
	/rpt_family="B2"
repeat_region	30612..30742
	/rpt_family="B4"
repeat_region	30745..30826
	/rpt_family="B4"
repeat_region	31880..31975
	/rpt_family="Alu"
repeat_region	33011..33224
	/rpt_family="MER1_type"
repeat_region	33885..33936
	/rpt_family="ERV1"
repeat_region	33996..34256
	/rpt_family="B4"
repeat_region	35781..35881
	/rpt_family="Alu"
repeat_region	36426..36633
	/rpt_family="B2"
repeat_region	36939..37054
	/rpt_family="L1"
repeat_region	37090..37214
	/rpt_family="L1"
repeat_region	37217..37808
	/rpt_family="L1"
repeat_region	38089..38236
	/rpt_family="Alu"
repeat_region	39404..39608
	/rpt_family="ERV1"
repeat_region	39609..40172
	/rpt_family="RMR15"
repeat_region	40205..40518

```

repeat_region      /rpt_family="ERV1"
                    40962. .41091
repeat_region      /rpt_family="Alu"
                    41573. .41676

Query Match       91.6%; Score 17.4; DB 9; Length 154904;
Best Local Similarity 94.7%; Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCCAAAATGA 19
    |||||
Db 111605 CCCAGTGCATCCCAAAATGA 111623

RESULT 10
AC069522/c
LOCUS          158620 bp      DNA      linear      HTG 06-SEP-2000
DEFINITION    Homo sapiens chromosome 3 clone RP11-129K1, WORKING DRAFT SEQUENCE,
                24 unordered pieces.
ACCESSION     AC069522
VERSION       AC069522.7 GI:99666643
KEYWORDS      HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE        Homo sapiens (human)
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
               Hominiidae; Homo.
REFERENCE     1 (bases 1 to 158620)
AUTHORS       Muzny, D.M., Adams, C., Bailey, M., Barbara, J., Blankenburg, K.,
               Bodocai, B., Bock, J., Bowie, S., Brooks, A., Buhay, C., Bunac, C.,
               Burdett, C., Burrows, J., Carter, M., Chacko, J., Chen, Z., Cox, C.,
               David, R., Delgado, O., Deshazo, D., Ding, Y., Domah-Rashid, N.,
               Dugan-Rocha, S., Durbin, K.J., Fernandez, C., Ferraguto, D.,
               Forcum-Taney, J., Frantz, P., Ganesh, R., Gorrell, J.H., Gorrell, L.L.,
               Guevara, W., Harris, K., Hernandez, J., Hodgson, A., Hogues, M.,
               Holloway, C., Hosak, H., Jackson, J.E., Jackson, L., Jia, Y., Jones, M.,
               Kelly, S., Kondrjewski, N., Kong, J., Kovar, C., Leal, B., Li, Z.,
               Lictarge, O., Liu, J., Liu, W., Logan, O., Lozado, R.J., Lu, J.,
               Lucier, R., Martin, R., Martinez, C., McLeod, M.P., Mei, G., Morgan, M.,
               Morris, S., Nash, S., Nelson, A., Nguyen, R., Nguyen, N., Nguyen, S.,
               Oswal, G., Parish, B., Paxton, S., Payton, B., Perez, L., Pu, L.L.,
               Quiles, M., Reiter, D., Rives, M., Samuel, S., Say, J., Scherer, S.,
               Shah, E., Shen, H., Simon, M., Sparks, A., Stamps, A., Sucgang, R.,
               Tabor, P., Taylor, T., Vasquez, L., Vinson, R., Vo, Q., Wabnah, M.,
               Watlington, S., Weinstein, G., Weinstein, I.R., Williamson, A.,
               Worley, K., Wren, J., Wrensford, G., Yu, W., Zhou, X., Nelson, D. and
               Gibbs, R.
               Direct Submission
               Unpublished
               2 (bases 1 to 158620)
               Worley, K.C.
               Direct Submission
               Submitted (02-JUN-2000) Human Genome Sequencing Center, Department
               of Molecular and Human Genetics, Baylor College of Medicine, One
               Baylor Plaza, Houston, TX 77030, USA
               On Sep 5, 2000 this sequence version replaced gi:9087381.
               ----- Genome Center
               Center: Baylor College of Medicine
               Center code: BCM
               Web site: http://www.hgsc.bcm.tmc.edu/
               Contact: hgsc-help@bcm.tmc.edu
               ----- Project Information
               Center project name: HBLU
               Center clone name: RP11-129K1
               ----- Summary Statistics
               Sequencing vector: M13; L08821
               Chemistry: Dye-primer Bodipy: 41% of reads
               Chemistry: Dye-terminator Big Dye: 58% of reads
               Assembly program: Phrap; version 0.990329
               Consensus quality: 143072 bases at least Q40
               Consensus quality: 149742 bases at least Q30
               Consensus quality: 152717 bases at least Q20
               Estimated insert size: 151232; sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 3.5x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 24768: contig of 24768 bp in length
* 24769: gap of unknown length
* 24868: contig of 18082 bp in length
* 42950: contig of unknown length
* 42951: gap of unknown length
* 43050: gap of unknown length
* 58519: contig of 15469 bp in length
* 58520: gap of unknown length
* 58619: contig of 10893 bp in length
* 69512: contig of unknown length
* 69513: gap of unknown length
* 69612: contig of 11452 bp in length
* 69613: gap of unknown length
* 81063: contig of 9549 bp in length
* 81065: gap of unknown length
* 81165: contig of unknown length
* 90713: gap of unknown length
* 90813: contig of 8795 bp in length
* 90814: gap of unknown length
* 99708: gap of unknown length
* 99709: contig of 6401 bp in length
* 106109: gap of unknown length
* 106209: contig of 8279 bp in length
* 108210: gap of unknown length
* 114488: contig of unknown length
* 114588: gap of unknown length
* 120453: contig of 5865 bp in length
* 120454: gap of unknown length
* 120553: contig of 3710 bp in length
* 124264: gap of unknown length
* 124363: contig of 5015 bp in length
* 124364: gap of unknown length
* 129378: contig of 5127 bp in length
* 129479: gap of unknown length
* 134605: contig of 5127 bp in length
* 134706: gap of unknown length
* 138119: contig of 3514 bp in length
* 138220: gap of unknown length
* 138320: contig of 3157 bp in length
* 141477: gap of unknown length
* 141576: contig of 3639 bp in length
* 145216: gap of unknown length
* 145315: contig of 2063 bp in length
* 145316: gap of unknown length
* 147378: contig of 1500 bp in length
* 147379: gap of unknown length
* 147478: contig of 2529 bp in length
* 148978: gap of unknown length
* 149078: contig of 2529 bp in length
* 151607: gap of unknown length
* 151608: contig of 1132 bp in length
* 152839: gap of unknown length
* 152840: contig of 1008 bp in length
* 152940: gap of unknown length
* 153948: contig of 1342 bp in length
* 154047: gap of unknown length
* 154048: contig of unknown length
* 155389: gap of unknown length
* 155489: contig of 1983 bp in length
* 155490: gap of unknown length
* 157472: contig of 1048 bp in length.
* 157473: gap of unknown length
* 157572: contig of 1048 bp in length.
* 157573: gap of unknown length
Location/Qualifiers
1..158620
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-129K1"
24769..24868
/estimated_length=unknown
42951..43050
/estimated_length=unknown
58520..58619
/estimated_length=unknown
FEATURES
source
gap
gap
gap

```

```
gap      69513..69612
/estimated_length=unknown
gap      81065..81164
/estimated_length=unknown
gap      90714..90813
/estimated_length=unknown
gap      99609..99708
/estimated_length=unknown
gap      106110..106209
/estimated_length=unknown
gap      114489..114588
/estimated_length=unknown
gap      120454..120553
/estimated_length=unknown
gap      124264..124363
/estimated_length=unknown
gap      129379..129478
/estimated_length=unknown
gap      134606..134705
/estimated_length=unknown
gap      138220..138319
/estimated_length=unknown
gap      141477..141576
/estimated_length=unknown
gap      145216..145315
/estimated_length=unknown
gap      147379..147478
/estimated_length=unknown
gap      148979..149078
/estimated_length=unknown
gap      151608..151707
/estimated_length=unknown
gap      152840..152939
/estimated_length=unknown
gap      153948..154047
/estimated_length=unknown
gap      155390..155489
/estimated_length=unknown
gap      157473..157572
/estimated_length=unknown

ORIGIN
Query Match      91.6%; Score 17.4; DB 14; Length 158620;
Best Local Similarity 94.7%; Pred.No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CCCAGTGCATCCAAATTGA 19
||||| ||||| ||||| ||||| |||||
Db      81044 CCCAGTTTCATCCAAATTGA 81026

RESULT 11
AC022376      160445 bp DNA linear HTG 02-MAR-2000
LOCUS      Homo sapiens chromosome 3p clone RP11-429p24, WORKING DRAFT
DEFINITION      SEQUENCE, 20 unordered pieces.
ACCESSION      AC022376
VERSION      AC022376.1 GI:6862624
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE      1 (bases 1 to 160445)
AUTHORS      Zhang,C., Lin,W., Dong,H., Wan,M., Xu,S., Gu,W., Tu,Y., Jia,J.,
Wu,C., Lu,G., Zhong,M., Zhou,Y., Ren,S., Fu,G., Chen,Z. and
Huang,M.
Direct Submission
TITLE      Submitted (03-FEB-2000) Genomic Dept., Chinese National Human
JOURNAL      Genome Center at Shanghai, 315 Guo Shou Jing Road, Shanghai,
Shanghai 201203, P. R. China
COMMENT      -----Genome Center Information-----
```

```
Center: Chinese National Human Genome Center at Shanghai
Center Code: CHGC Web site: http://www.chgc.sh.cn
Email: mhuang@chgc.sh.cn or fugang@chgc.sh.cn
-----End Genome Center Information-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*      1      2329: contig of 2329 bp in length
*      2330      gap of unknown length
*      5118: contig of 2789 bp in length
*      5119      gap of unknown length
*      8982: contig of 3864 bp in length
*      8983      gap of unknown length
*      13408: contig of 4426 bp in length
*      13409      gap of unknown length
*      17836: contig of 4428 bp in length
*      17837      gap of unknown length
*      21881: contig of 4045 bp in length
*      21882      gap of unknown length
*      26608: contig of 4727 bp in length
*      26609      gap of unknown length
*      32708: contig of 6100 bp in length
*      32709      gap of unknown length
*      40332: contig of 7624 bp in length
*      40333      gap of unknown length
*      48645: contig of 8313 bp in length
*      48646      gap of unknown length
*      55410: contig of 6765 bp in length
*      55411      gap of unknown length
*      62714: contig of 7304 bp in length
*      62715      gap of unknown length
*      72977: contig of 10263 bp in length
*      72978      gap of unknown length
*      82001: contig of 9024 bp in length
*      82002      gap of unknown length
*      93484: contig of 11483 bp in length
*      93485      gap of unknown length
*      104097: contig of 10613 bp in length
*      104098      gap of unknown length
*      114643: contig of 10546 bp in length
*      114644      gap of unknown length
*      130461: contig of 15818 bp in length
*      130462      gap of unknown length
*      141784: contig of 11323 bp in length
*      141785      gap of unknown length
*      160445: contig of 18661 bp in length.

FEATURES
            Location/Qualifiers
            1..160445
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="3p"
            /clone="RP11-429p24"

ORIGIN
Query Match      91.6%; Score 17.4; DB 14; Length 160445;
Best Local Similarity 94.7%; Pred.No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CCCAGTGCATCCAAATTGA 19
||||| ||||| ||||| ||||| |||||
Db      158131 CCCAGTTTCATCCAAATTGA 158149

RESULT 12
AC016963/c
LOCUS      AC016963
DEFINITION      Homo sapiens chromosome 3 clone RP11-121D3, WORKING DRAFT SEQUENCE,
```

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

16 unordered pieces.  
AC016963  
AC016963.10 GI:9719649  
HTG; HTGS PHASE1; HTGS\_DRAFT.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominoidea; Homo.  
1 (bases 1 to 160722)  
Muzny, D.M., Adams, C., Bailey, M., Barbara, J., Blankenburg, K.,  
Bodurka, B., Bouck, J., Bowie, S., Brooks, A., Buhay, C., Bunick, C.,  
Burkett, C., Burrows, J., Carter, M., Chacko, J., Chen, Z., Cox, C.,  
David, R., Delgado, O., Deshazo, D., Ding, Y., Domah-Rashid, N.,  
Dugan-Rocha, S., Durbin, K.J., Fernandez, C., Ferraguto, D.,  
Forcum-Taney, J., Frantz, P., Ganes, R., Gorrell, J.H., Gorrell, L.L.,  
Guevara, M., Harris, K., Hernandez, J., Hodgson, A., Hogue, M.,  
Holloway, C., Hosak, H., Jackson, L.E., Jackson, L., Jia, Y., Jones, M.,  
Kelly, S., Kondrjewski, N., Kong, Y., Kovar, C., Leal, B., Li, Z.,  
Lichter, O., Liu, J., Liu, W., Logan, O., Lozano, R.J., Lu, J.,  
Lucier, R., Martin, R., Martinez, C., McLeod, M.P., Mei, G., Morgan, M.,  
Morris, S., Nash, S., Nelson, A., Nguyen, R., Nguyen, N., Nguyen, S.,  
Oswal, G., Parish, B., Paxton, S., Payton, B., Perez, L., Pu, L.L.,  
Quiles, M., Reiter, D., Rives, M., Samuel, S., Say, J., Scherer, S.,  
Shah, E., Shen, H., Simon, M., Sparks, A., Stamps, A., Sucgang, R.,  
Tabor, P., Taylor, T., Vasquez, L., Vinson, R., Vo, Q., Wabwah, M.,  
Watlington, S., Weinstein, G., Weinstein, I.R., Williamson, A.,  
Worley, K., Wren, J., Wrenford, G., Yu, W., Zhou, X., Nelson, D. and  
Gibbs, R.

REFERENCE  
AUTHORS

Direct Submission  
Unpublished  
2 (bases 1 to 160722)  
Worley, K.C.  
Direct Submission  
Submitted (09-DEC-1999) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Aug 7, 2000 this sequence version replaced gi:8493478.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: HMVI  
Center clone name: RP11-121D3  
----- Summary Statistics  
Sequencing vector: M13; L08821  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 137317 bases at least Q40  
Consensus quality: 148491 bases at least Q30  
Consensus quality: 154780 bases at least Q20  
Estimated insert size: 154776; sum-of-contigs estimation  
Estimated insert size: 167977; agarose-fp estimation  
Quality coverage: 2.8x in Q20 bases; agarose-fp estimation  
Quality coverage: 3x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
consists of 16 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.  
\* 1 31191: contig of 31191 bp in length  
\* 31192: gap of unknown length  
\* 31292: contig of 26029 bp in length  
\* 57321: gap of unknown length  
\* 57421: contig of 17069 bp in length

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

1. 160722  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="3"  
/clone="RP11-121D3"  
31192..31291  
/estimated\_length=unknown  
57321..57420  
/estimated\_length=unknown  
74490..74589  
/estimated\_length=unknown  
87084..87183  
/estimated\_length=unknown  
97246..97345  
/estimated\_length=unknown  
109229..109328  
/estimated\_length=unknown  
117749..117848  
/estimated\_length=unknown  
127378..127477  
/estimated\_length=unknown  
135762..135861  
/estimated\_length=unknown  
144345..144444  
/estimated\_length=unknown  
149052..149151  
/estimated\_length=unknown  
153814..153913  
/estimated\_length=unknown  
156691..156790  
/estimated\_length=unknown  
158351..158450  
/estimated\_length=unknown  
159508..159607  
/estimated\_length=unknown

COMMENT

Location/Qualifiers  
1. 160722  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="3"  
/clone="RP11-121D3"  
31192..31291  
/estimated\_length=unknown  
57321..57420  
/estimated\_length=unknown  
74490..74589  
/estimated\_length=unknown  
87084..87183  
/estimated\_length=unknown  
97246..97345  
/estimated\_length=unknown  
109229..109328  
/estimated\_length=unknown  
117749..117848  
/estimated\_length=unknown  
127378..127477  
/estimated\_length=unknown  
135762..135861  
/estimated\_length=unknown  
144345..144444  
/estimated\_length=unknown  
149052..149151  
/estimated\_length=unknown  
153814..153913  
/estimated\_length=unknown  
156691..156790  
/estimated\_length=unknown  
158351..158450  
/estimated\_length=unknown  
159508..159607  
/estimated\_length=unknown

ORIGIN

Query Match 91.6%; Score 17.4; DB 14; Length 160722;  
Best Local Similarity 94.7%; Pred.No. 1.6e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CCCAGTGCATCCCAATTGA 19

Db 66821 CCCAGTTCATCCCAATTGA 66803

```

RESULT 13
AC026882/c
LOCUS
DEFINITION
Homo sapiens chromosome 3 clone RP11-211K13 map 3p, complete
sequence.
AC026882
AC026882.6 GI:24080640
VERSION
HTG; HTGS_DRAFT.
KEYWORDS
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
AUTHORS
Xiong,H., Zhou,Y., Dong,H., Lin,W., Chen,B., Zhang,C., Zhang,Y.,
Cai,Z., Ying,H.F., Wang,H., Gu,W., Zhu,G., Tu,Y., Zhang,X., Jia,J.,
Shen,H., Zhang,D., Wu,C., Lu,G., Zhong,M., Jiang,H., Ren,S., Fu,G.,
Chen,Z. and Huang,M.
1 (bases 1 to 162422)
TITLE
Chromosome 3p genomic sequence
JOURNAL
Unpublished
REFERENCE
AUTHORS
Li,W., Hu,S., Dong,W., Wang,J., Zhang,Y., Zhang,H., Liu,B., Bao,W.,
Sun,Y., Wu,Q., Wang,H., Yang,X., Cheng,C., Wang,Y., Niu,Y., Qi,X.,
Liu,T., Zhang,H., Liu,N., Wu,D., Yu,B., Fan,H., Liu,Y., Li,G.,
Li,C., Bao,Q., Bao,J., Wang,X., Song,L., Zhang,L., Guo,D.,
Huang,F., Zhang,G., Li,J., Bian,X., Zhang,M., Li,L., Feng,X., Yu,J.
and Yang,H.
1 (bases 1 to 162422)
2 (bases 1 to 162422)
TITLE
Direct Submission
JOURNAL
Submitted (25-MAR-2000) Human Genomic Center, Institute of
Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
100101, P.R.China
REFERENCE
AUTHORS
Xiong,H., Zhou,Y., Dong,H., Lin,W., Chen,B., Zhang,C., Zhang,Y.,
Cai,Z., Ying,H.F., Wang,H., Gu,W., Zhu,G., Tu,Y., Zhang,X., Jia,J.,
Shen,H., Zhang,D., Wu,C., Lu,G., Zhong,M., Jiang,H., Ren,S., Fu,G.,
Chen,Z. and Huang,M.
3 (bases 1 to 162422)
TITLE
Direct Submission
JOURNAL
Submitted (05-JAN-2001) 1. Chinese Human Genome Center at Shanghai
2. Institute of Genetics, Chinese Academy of Sciences, P.R.China
4 (bases 1 to 162422)
REFERENCE
AUTHORS
Bao,J., Bao,Q., Bao,W., Bian,X., Cao,T., Chen,C., Chen,J., Ding,H.,
Dong,W., Fan,H., Feng,X., Gong,J., Guan,Q., Gu,X., Guo,D., Guo,Z.,
He,L., Hu,S., Huang,F., Jin,Y., Kang,N., Li,C., Li,C., Li,F.,
Li,G., Li,J., Li,L., Li,S., Li,T., Liu,Y., Liu,N., Liu,B., Liu,Y.,
Li,W., Li,W., Li,X., Luo,C., Luo,J., Niu,Y., Qi,Q., Qi,X., Song,L.,
Song,S., Sun,M., Sun,W., Sun,Y., Tan,X., Tao,R., Wang,H., Wang,J.,
Wang,Y., Wang,J., Wang,L., Wang,L., Wang,R., Wang,X., Wang,X.,
Wang,Y., Wu,D., Wu,Q., Xie,F., Xuan,Z., Xue,Y., Yan,C., Yang,X.,
Yu,B., Zeng,Y., Zhang,G., Zhang,H., Zhang,H., Zhang,L., Zhang,M.,
Zhang,X., Zhang,X., Zhang,Y., Zhang,Y., Zhang,Z., Zhu,B., Zhu,N.,
Yu,J. and Yang,H.
1 (bases 1 to 162422)
TITLE
Direct Submission
JOURNAL
Submitted (11-FEB-2001) Human Genomic Center, Institute of
Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
100101, P.R.China
5 (bases 1 to 162422)
REFERENCE
AUTHORS
Bao,J., Bao,Q., Bao,W., Bian,X., Cao,T., Chen,C., Chen,J., Ding,H.,
Dong,W., Fan,H., Feng,X., Gong,J., Guan,Q., Gu,X., Guo,D., Guo,Z.,
He,L., Hu,S., Huang,F., Jin,Y., Kang,N., Li,C., Li,C., Li,F.,
Li,G., Li,J., Li,L., Li,S., Li,T., Liu,Y., Liu,N., Liu,B., Liu,Y.,
Li,W., Li,W., Li,X., Luo,C., Luo,J., Niu,Y., Qi,Q., Qi,X., Song,L.,
Song,S., Sun,M., Sun,W., Sun,Y., Tan,X., Tao,R., Wang,H., Wang,J.,
Wang,Y., Wang,J., Wang,L., Wang,L., Wang,R., Wang,X., Wang,X.,
Wang,Y., Wu,D., Wu,Q., Xie,F., Xuan,Z., Xue,Y., Yan,C., Yang,X.,
Yu,B., Zeng,Y., Zhang,G., Zhang,H., Zhang,H., Zhang,L., Zhang,M.,
Zhang,X., Zhang,X., Zhang,Y., Zhang,Y., Zhang,Z., Zhu,B., Zhu,N.,
Yu,J. and Yang,H.
1 (bases 1 to 162422)
TITLE
Direct Submission
JOURNAL
Submitted (17-OCT-2002) Human Genomic Center, Institute of
Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
100101, P.R.China
On Oct 17, 2002 this sequence version replaced gi:12745074.

```

```

-----Genome Center
Center:Beijing Center
Center code:Beijing
Website:http://hgsc.igtp.ac.cn
http://www.genomics.org.cn
Contact:hgsc@igtp.ac.cn
----- Project Information
Center project name:1k project
Center clone name: RP11-211K13
----- Summary Statistics
Sequencing vector: pUC18; 100% of reads
Chemistry: Dye-terminator Big Dye; 45% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 165451 bases at least Q40
Consensus quality: 165592 bases at least Q30
Consensus quality: 165628 bases at least Q20
Insert size: 162422; sum-of-contigs
Quality coverage: 9.09x in Q20 bases;sum-of-contigs
-----
Location/Qualifiers
1. .162422
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3"
/map="3p"
/clone="RP11-211K13"

FEATURES
source
Query Match 91.6%; Score 17.4; DB 8; Length 162422;
Best Local Similarity 94.7%; Pred No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCACGTGCATCCAAATTGA 19
|||||
Db 5720 CCACGTGCATCCAAATTGA 5702

RESULT 14
BX530079/c
LOCUS
DEFINITION
Zebrafish DNA sequence from clone DKEY-51A16 in linkage group 13,
complete sequence.
ACCESSION
BX530079
VERSION
BX530079.19 GI:53145127
KEYWORDS
HTG.
SOURCE
Danio rerio (zebrafish)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 166397)
Direct Submission
Donaldson,S.
TITLE
Direct Submission
JOURNAL
Submitted (30-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zf1sh-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Sep 30, 2004 this sequence version replaced gi:52839527.
COMMENT
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zf1sh-help@sanger.ac.uk
-----
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate

```



chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) Clone-derived Zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.

Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhirong Bao and Sean Eddy, submitted), and those beginning 'drir' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see [http://www.sanger.ac.uk/Projects/D\\_rexio/fishmask.shtml](http://www.sanger.ac.uk/Projects/D_rexio/fishmask.shtml) DKEY-51A16 is from a Zebrafish BAC library

VECTOR: pIndigoBAC-5.

#### FEATURES

source  
1..166397  
Location/Qualifiers  
/organism="Danio rerio"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7955"  
/clone="DKEY-51A16"  
/clone\_lib="DanioKey"

#### ORIGIN

Query Match 91.6%; Score 17.4; DB 5; Length 166397;  
Best Local Similarity 94.7%; Pred. No. 1.6e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CCCAGTGCATCCAAATTGA 19

Db 103995 CCCAGGGCATCCAAATTGA 103977

#### RESULT 15

EX321918/c

LOCUS BX321918 170084 bp DNA linear HTG 10-AUG-2004  
DEFINITION Danio rerio clone CH211-230G11, WORKING DRAFT SEQUENCE, 2 unordered pieces.

#### ACCESSION

EX321918

VERSION BX321918.6 GI:51100631

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_ACTIVEFIN; HTGS\_DRAFT; HTGS\_FULLTOP.

SOURCE Danio rerio (zebrafish)

#### ORGANISM

Danio rerio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
Cypriniformes; Cyprinidae; Danio.

#### REFERENCE

AUTHORS Kay,M.

TITLE Direct Submission

JOURNAL Submitted (09-AUG-2004) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: [zf1sh-help@sanger.ac.uk](mailto:zf1sh-help@sanger.ac.uk)  
zf1sh-help@sanger.ac.uk Clone requests: [clonerequest@sanger.ac.uk](mailto:clonerequest@sanger.ac.uk)  
On Aug 10, 2004 this sequence version replaced gi:32398461.

#### COMMENT

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: [zf1sh-help@sanger.ac.uk](mailto:zf1sh-help@sanger.ac.uk)

----- Project Information

Center project name: zC230G11

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Chemistry: Dye-terminator; 100% of reads

Consensus quality: 169798 bases at least Q40

Consensus quality: 169867 bases at least Q30

Consensus quality: 169918 bases at least Q20

Insert size: 169984; sum-of-contigs

Insert size: 180071; 17.4% error; agarose-fp

Quality coverage: 6.66x in Q20 bases; sum-of-contigs Quality

coverage: 6.82x in Q20 bases; agarose-fp

-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 2 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 122253: contig of 122253 bp in length

\* 122254 122353: gap of 100 bp

\* 122354 170084: contig of 47731 bp in length.

#### FEATURES

source  
1..170084  
Location/Qualifiers  
/organism="Danio rerio"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7955"  
/clone="CH211-230G11"  
/clone\_lib="CHORI-211"

#### misc\_feature

1..122253  
/note="assembly fragment:01029  
fragment chain:1"

#### misc\_feature

122354..170084  
/note="assembly fragment:00675  
fragment chain:1"

#### ORIGIN

Query Match 91.6%; Score 17.4; DB 14; Length 170084;  
Best Local Similarity 94.7%; Pred. No. 1.6e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CCCAGTGCATCCAAATTGA 19

Db 110167 CCCAGGGCATCCAAATTGA 110149

#### RESULT 16

AC018498

LOCUS AC018498 172111 bp DNA linear HTG 03-FEB-2000  
DEFINITION Homo sapiens chromosome 3p clone RP11-263J8, WORKING DRAFT SEQUENCE, 7 unordered pieces.

#### ACCESSION

AC018498

VERSION AC018498.3 GI:6684198

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE Homo sapiens (human)

#### ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominoidea; Homo.

#### REFERENCE

AUTHORS

1 (bases 1 to 172111)  
Bao,Q., Hu,S., Dong,W., Zhang,X., Wang,J., Zhang,Y., Zhang,H.,  
Liu,B., Bao,W., Sun,Y., Wu,Q., Wang,H., Yang,X., Cheng,C., Wang,Y.,  
Niu,Y., Qi,X., Li,T., Zhang,H., Liu,N., Wu,D., Yu,B., Fan,H.,  
Liu,Y., Li,G., Li,C., Wang,X., Bao,J., Wang,X., Song,L., Zhang,L.,  
Guo,D., Huang,F., Zhang,G., Li,J., Bian,X., Zhang,M., Li,D.,  
Feng,X., Yu,J. and Yang,H.

#### TITLE

Chromosome 3p genomic sequence

#### JOURNAL

REFERENCE

AUTHORS

2 (bases 1 to 172111)  
Wu,Q., Hu,S., Dong,W., Wang,J., Zhang,Y., Zhang,H., Liu,B., Bao,W.,  
Sun,Y., Wang,X., Wang,H., Yang,X., Cheng,C., Wang,Y., Niu,Y.,  
Qi,X., Li,T., Zhang,H., Liu,N., Wu,D., Yu,B., Fan,H., Liu,Y.,  
Li,G., Li,C., Bao,Q., Bao,J., Wang,X., Song,L., Zhang,L., Guo,D.,  
Huang,F., Zhang,G., Li,J., Bian,X., Zhang,M., Li,L., Feng,X., Yu,J.  
and Yang,H.

#### TITLE

Direct Submission

Submitted (13-DEC-1999) Human Genomic Center, Institute of  
Genetics, Chinese Academy of Sciences, Datun Road, Beijing,  
100101, P.R.China

COMMENT	On Jan 9, 2000 this sequence version replaced gi:6633937. * NOTE: This is a 'working draft' sequence. It currently consists of 7 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.	JOURNAL
FEATURES	1 4116: contig of 4116 bp in length 1 4117 8949: contig of 4833 bp in length 1 8950 19559: contig of 10610 bp in length 1 19560 33335: contig of 13776 bp in length 1 33336 57232: contig of 23897 bp in length 1 57233 108094: contig of 50862 bp in length 1 108095 172111: contig of 64017 bp in length. Location/Qualifiers 1. .172111 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="3p" /clone="RP11-263J8"	REFERENCE AUTHORS
ORIGIN	Query Match 91.6%; Score 17.4; DB 14; Length 172111; Best Local Similarity 94.7%; Pred. No. 1.6e+02; Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  Qy 1 CCCAGTGCATCCCAAAATTGA 19                       Db 36845 CCCAGTTCATCCCAAAATTGA 36863	REFERENCE AUTHORS
RESULT 17	AC024057/c 186007 bp DNA linear PRI 25-NOV-2002	JOURNAL
LOCUS	Homo sapiens chromosome 3 clone RP11-299N3 map 3p, complete sequence.	REFERENCE AUTHORS
ACCESSION	AC024057 AC087187	REFERENCE AUTHORS
VERSION	AC024057.5 GI:25281353	REFERENCE AUTHORS
KEYWORDS	HTG; HTGS DRAFT.	REFERENCE AUTHORS
SOURCE	Homo sapiens (human)	REFERENCE AUTHORS
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo. 1 (bases 1 to 186007) Xiong, H., Zhou, Y., Dong, H., Lin, W., Chen, B., Zhang, C., Zhang, Y., Cai, Z., Ying, H. F., Wang, H., Gu, W., Zhu, G., Tu, Y., Zhang, X., Jia, J., Shen, H., Zhang, D., Wu, C., Lu, G., Zhong, M., Jiang, H., Ren, S., Fu, G., Chen, Z. and Huang, M.	REFERENCE AUTHORS
REFERENCE	1 (bases 1 to 186007) Xiong, H., Zhou, Y., Dong, H., Lin, W., Chen, B., Zhang, C., Zhang, Y., Cai, Z., Ying, H. F., Wang, H., Gu, W., Zhu, G., Tu, Y., Zhang, X., Jia, J., Shen, H., Zhang, D., Wu, C., Lu, G., Zhong, M., Jiang, H., Ren, S., Fu, G., Chen, Z. and Huang, M.	REFERENCE AUTHORS
TITLE	Chromosome 3p genomic sequence	REFERENCE AUTHORS
JOURNAL	Unpublished	REFERENCE AUTHORS
REFERENCE	2 (bases 1 to 186007) Wang, H., Zhao, Y., Zhang, C., Lin, W., Dong, H., Wan, M., Xu, S., Gu, W., Tu, Y., Jia, J., Wu, C., Lu, G., Zhong, M., Zhou, Y., Ren, S., Fu, G., Chen, Z. and Huang, M.	REFERENCE AUTHORS
TITLE	Direct Submission	REFERENCE AUTHORS
JOURNAL	Genome Center at Shanghai, 315 Guo Shou Jing Road, Shanghai, Shanghai 201203, P. R. China	REFERENCE AUTHORS
REFERENCE	3 (bases 1 to 186007) Zhou, Y., Lin, W., Zhang, C., Dong, H., Wan, M., Xu, S., Gu, W., Tu, Y., Jia, J., Wu, C., Lu, G., Zhong, M., Ren, S., Xiong, H., Fu, G., Chen, Z. and Huang, M.	REFERENCE AUTHORS
TITLE	Direct Submission	REFERENCE AUTHORS

Submitted (24-AUG-2000) Genomic Dept., Chinese National Human Genome Center at Shanghai, 315 Guo Shou Jing Road, Shanghai, Shanghai 201203, P. R. China	JOURNAL
4 (bases 1 to 186007) Zhou, Y., Xiong, H., Dong, H., Lin, W., Chen, B., Zhang, C., Zhang, Y., Cai, Z., Ying, H. F., Wang, H., Gu, W., Zhu, G., Tu, Y., Zhang, X., Jia, J., Shen, H., Zhang, D., Wu, C., Lu, G., Zhong, M., Jiang, H., Ren, S., Fu, G., Chen, Z. and Huang, M.	REFERENCE AUTHORS
Direct Submission	REFERENCE AUTHORS
Submitted (04-JAN-2001) 1. Chinese Human Genome Center at Shanghai	REFERENCE AUTHORS
2. Institute of Genetics, Chinese Academy of Sciences, P. R. China	REFERENCE AUTHORS
5 (bases 1 to 186007) Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H., Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D., Guo, Z., He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C., Li, F., Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, B., Liu, Y., Li, W., Li, W., Li, Y., Luo, C., Luo, J., Niu, Y., Qi, Q., Qi, X., Song, L., Li, W., Li, W., Li, Y., Luo, C., Luo, J., Niu, Y., Qi, Q., Qi, X., Song, L., Song, S., Sun, M., Sun, W., Sun, Y., Tan, X., Tao, R., Wang, H., Wang, J., Wang, J., Wang, L., Wang, L., Wang, R., Wang, X., Wang, X., Wang, Y., Wang, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X., Zeng, Y., Zhang, G., Zhang, H., Zhang, H., Zhang, L., Zhang, M., Zhang, X., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N., Yu, J. and Yang, H.	REFERENCE AUTHORS
Direct Submission	REFERENCE AUTHORS
Submitted (15-FEB-2001) Human Genomic Center, Institute of Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing 100101, P. R. China	REFERENCE AUTHORS
6 (bases 1 to 186007) Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H., Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D., Guo, Z., He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C., Li, F., Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, B., Liu, Y., Li, W., Li, W., Li, Y., Luo, J., Niu, Y., Qi, Q., Qi, X., Song, L., Song, S., Sun, M., Sun, W., Sun, Y., Tan, X., Tao, R., Wang, H., Wang, J., Wang, J., Wang, L., Wang, L., Wang, R., Wang, X., Wang, X., Wang, Y., Wang, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X., Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, H., Zhang, L., Zhang, M., Zhang, X., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N., Yu, J. and Yang, H.	REFERENCE AUTHORS
Direct Submission	REFERENCE AUTHORS
Submitted (11-APR-2002) Human Genomic Center, Institute of Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing 100101, P. R. China	REFERENCE AUTHORS
7 (bases 1 to 186007) Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H., Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D., Guo, Z., He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C., Li, F., Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, B., Liu, Y., Li, W., Li, W., Li, Y., Luo, C., Luo, J., Niu, Y., Qi, Q., Qi, X., Song, L., Song, S., Sun, M., Sun, W., Sun, Y., Tan, X., Tao, R., Wang, H., Wang, J., Wang, J., Wang, L., Wang, L., Wang, R., Wang, X., Wang, X., Wang, Y., Wang, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X., Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, H., Zhang, L., Zhang, M., Zhang, X., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N., Yu, J. and Yang, H.	REFERENCE AUTHORS
Direct Submission	REFERENCE AUTHORS
Submitted (11-APR-2002) Human Genomic Center, Institute of Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing 100101, P. R. China	REFERENCE AUTHORS
8 (bases 1 to 186007) Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H., Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D., Guo, Z., He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C., Li, F., Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, B., Liu, Y., Li, W., Li, W., Li, Y., Luo, C., Luo, J., Niu, Y., Qi, Q., Qi, X., Song, L., Song, S., Sun, M., Sun, W., Sun, Y., Tan, X., Tao, R., Wang, H., Wang, J., Wang, J., Wang, L., Wang, L., Wang, R., Wang, X., Wang, X., Wang, Y., Wang, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X., Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, H., Zhang, L., Zhang, M., Zhang, X., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N., Yu, J. and Yang, H.	REFERENCE AUTHORS
Direct Submission	REFERENCE AUTHORS
Submitted (25-NOV-2002) Human Genomic Center, Institute of Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing 100101, P. R. China	REFERENCE AUTHORS
On Nov 25, 2002 this sequence version replaced gi:12831358.	REFERENCE AUTHORS
Center:Beijing Center	REFERENCE AUTHORS
Center code:Beijing	REFERENCE AUTHORS
Website:http://hgsc.igtp.ac.cn	REFERENCE AUTHORS
Contact:hgsc@igtp.ac.cn	REFERENCE AUTHORS
Project Information	REFERENCE AUTHORS
Center project name:1% project	REFERENCE AUTHORS
Center clone name: RP11-299N3	REFERENCE AUTHORS
Summary Statistics	REFERENCE AUTHORS
Sequencing vector: pUC18; 100% of reads	REFERENCE AUTHORS
Chemistry: Dye-terminator; ET 55% of reads	REFERENCE AUTHORS
Chemistry: Dye-terminator Big Dye 45% of reads	REFERENCE AUTHORS

Submitted (24-AUG-2000) Genomic Dept., Chinese National Human Genome Center at Shanghai, 315 Guo Shou Jing Road, Shanghai, Shanghai 201203, P. R. China	
4 (bases 1 to 186007)	
Zhou, Y., Xiong, H., Dong, H., Lin, W., Chen, B., Zhang, C., Zhang, Y., Cai, Z., Ying, H. F., Wang, H., Gu, W., Zhu, G., Tu, Y., Zhang, X., Jia, J., Shen, H., Zhang, D., Wu, C., Lu, G., Zhong, M., Jiang, H., Ren, S., Fu, G., Chen, Z. and Huang, M.	
Direct Submission	
Submitted (04-JAN-2001) 1. Chinese Human Genome Center at Shanghai	
2. Institute of Genetics, Chinese Academy of Sciences, P.R.China	
5 (bases 1 to 186007)	
Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H., Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D., Guo, Z., He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C., Li, F., Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, B., Liu, Y., Li, W., Li, W., Li, Y., Luo, C., Luo, J., Niu, Y., Qi, Q., Qi, X., Song, L., Song, S., Sun, M., Sun, W., Sun, Y., Tan, X., Tao, R., Wang, H., Wang, J., Wang, J., Wang, L., Wang, L., Wang, R., Wang, X., Wang, X., Wang, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X., Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, H., Zhang, L., Zhang, M., Zhang, X., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N., Yu, J. and Yang, H.	
Direct Submission	
Submitted (15-FEB-2001) Human Genomic Center, Institute of Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing 100101, P.R.China	
6 (bases 1 to 186007)	
Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H., Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D., Guo, Z., He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C., Li, F., Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, B., Liu, Y., Li, W., Li, W., Li, Y., Luo, J., Niu, Y., Qi, Q., Qi, X., Song, L., Song, S., Sun, M., Sun, W., Sun, Y., Tan, X., Tao, R., Wang, H., Wang, J., Wang, J., Wang, L., Wang, L., Wang, R., Wang, X., Wang, X., Wang, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X., Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, H., Zhang, L., Zhang, M., Zhang, X., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N., Yu, J. and Yang, H.	
Direct Submission	
Submitted (11-APR-2002) Human Genomic Center, Institute of Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing 100101, P.R.China	
7 (bases 1 to 186007)	
Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H., Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D., Guo, Z., He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C., Li, F., Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, B., Liu, Y., Li, W., Li, W., Li, Y., Luo, C., Luo, J., Niu, Y., Qi, Q., Qi, X., Song, L., Song, S., Sun, M., Sun, W., Sun, Y., Tan, X., Tao, R., Wang, H., Wang, J., Wang, J., Wang, L., Wang, L., Wang, R., Wang, X., Wang, X., Wang, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X., Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, H., Zhang, L., Zhang, M., Zhang, X., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N., Yu, J. and Yang, H.	
Direct Submission	
Submitted (25-NOV-2002) Human Genomic Center, Institute of Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing 100101, P.R.China	
On Nov 25, 2002 this sequence version replaced gi:12831358.	
-----Genome Center	
Center:Beijing Center	
Center code:Beijing	
Website:http://hgsc.igtp.ac.cn	
http://www.genomics.org.cn	
Contact:hgc@igtp.ac.cn	
----- Project Information	
Center project name:l% project	
Center Clone name: RP11-299N3	
----- Summary Statistics	
Sequencing vector: pUC18; 100% of reads	
Chemistry: Dye-terminator; ET 55% of reads	
Chemistry: Dye-terminator Big Dye; 45% of reads	

Quality coverage: 1.58x in Q20 bases; sum-of-contigs

```

/clone="RP11-299N3"

```

Best Local Similarity 94.7%; Pred. No. 1.6e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Dbb 132470 CCCAGTTTCATCCAAATTGA 132452

sequence.

**SOURCE** Mus musculus (house mouse)

**AUTHORS**  
Hopkins, B.

cambridge@cam.ac.uk; 2 mail enquiries:  
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

----- Genome Center

contact: [namqaezj@baidu.com](mailto:namqaezj@baidu.com)  
-----

Only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all

The following abbreviations are used to associate primary accession

PROBATION DEPARTMENT, 1470, UNIVERSITY AVENUE, ALBANY, NEW YORK

Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umanai, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, J., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, X., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

## TITLE

Direct Submission

## JOURNAL

Unpublished

## AUTHORS

2 (bases 1 to 212916)

## JOURNAL

Worley, K.C.

## AUTHORS

Submitted (03-OCT-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 212916)

## REFERENCE

Rat Genome Sequencing Consortium.

## JOURNAL

Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On May 10, 2003 this sequence version replaced gi:22855336.

## COMMENT

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information

Center project name: GIDC

Center clone name: CH230-192P2

----- Summary Statistics

Assembly program: Atlas 3.0;

Consensus quality: 207247 bases at least Q40

Consensus quality: 208474 bases at least Q30

Consensus quality: 209016 bases at least Q20

Estimated insert size: 214383; sum-of-contigs estimation

Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 2 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 211596: contig of 211596 bp in length

\* 211697 211696: gap of unknown length

\* 211697 212916: contig of 1220 bp in length.

\* Location/Qualifiers

1. .212916

## FEATURES

source

```

/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/dn_xref="taxon:10116"
/clone="CH230-192P2"
205875..208050
misc_feature
/note="wgs contig"
208934..211596
misc_feature
/note="wgs contig"
211597..211696
/estimated_length=unknown

ORIGIN
Query Match          91.6%; Score 17.4; DB 14; Length 212916;
Best Local Similarity 94.7%; Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCCAGTCATCCAAATTGA 19
   |||||
Db 40785 CCCAGTCCTCCAAATTGA 40803

RESULT 20
AC135524
LOCUS
DEFINITION
Rattus norvegicus clone CH230-225E21, WORKING DRAFT SEQUENCE, 2
unordered pieces.
AC135524
AC135524.2 GI:25073237
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS
Rattus norvegicus (Norway rat)
SOURCE
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 226412)
REFERENCE
Muzny, D., Marie, H., Metzker, M., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Bandwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleaveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, P., Evans, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kuwis, C., Kraft, C., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshewari, L., Louisedge, H., Lozado, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Naif, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwaokemele, O., Okwuon, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,

```

Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaseca, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wood, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

## TITLE

## JOURNAL

Unpublished

2 (bases 1 to 226412)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (18-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 226412)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

On Nov 19, 2002 this sequence version replaced gi:24110922. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information

Center project name: KCOE

Center clone name: CH230-225E21

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 218594 bases at least Q40

Consensus quality: 220639 bases at least Q30

Consensus quality: 221575 bases at least Q20

Estimated insert size: 224701; sum-of-contigs estimation

Quality coverage: 9x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 225094: contig of 225094 bp in length

\* 225095 225194: gap of unknown length

\* 225195 226412: contig of 1218 bp in length.

Location/Qualifiers

1. .226412

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clone="CH230-225E21"

## FEATURES

## source

misc\_feature 13..888  
/note="clone boundary  
clone end:T7  
site:  
end sequence:B2101274"  
complement(224142..224997)  
/note="clone boundary  
clone end:Sp6  
site:  
end sequence:B2101274"  
225095..225194  
/estimated\_length=unknown

gap

ORIGIN

Query Match 91.6%; Score 17.4; DB 14; Length 226412;  
Best Local Similarity 94.7%; Pred. No. 1.6e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATTGA 19  
|||||  
Db 33866 CCCAGTGCCTCCAAATTGA 33884  
|||||

RESULT 21  
AC134630/c

LOCUS  
AC134630 Rattus norvegicus clone CH230-77J24, \*\*\* SEQUENCING IN PROGRESS  
DEFINITION  
\*\*\*, 2 unordered pieces.  
ACCESSION  
AC134630 GI:25138865  
VERSION  
HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
KEYWORDS  
Rattus norvegicus (Norway rat)  
SOURCE  
Rattus norvegicus  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridea; Muridae; Murinae; Rattus.  
REFERENCE  
1 (bases 1 to 237447)  
Muzny, D., Maric, M., Metzker, M., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, M., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Hernandez, R., Hines, P., Hawes, A., Henderson, N., Hernandez, J., Harvey, Y., Havlak, P., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, D., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Loresuhewa, L., Loulseged, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Naif, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwackemele, O., Okwono, G., Olarnpunaagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Flopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Fuafo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,

Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,  
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,  
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D.,  
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,  
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,  
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,  
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,  
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
Williams, G., Willison, R., Wlezyk, R., Wooden, H., Worley, K.,  
Wright, D., Wright, J., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,  
Weinstock, G. and Gibbs, R.A.

Direct Submission  
Unpublished  
2 (bases 1 to 237447)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (28-SEP-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 237447)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

On Nov 20, 2002 this sequence version replaced gi:23343613.  
The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
in the feature table below represents a scaffold in the Atlas  
assembly (a 'contig-scaffold'). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented, and separated  
by sized gaps filled with Ns to the estimated size. The sequence  
may extend beyond the ends of the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole  
genome shotgun sequence reads. Both end sequences and whole genome  
shotgun sequence only contigs will be indicated in the feature  
table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: KCLX  
Center clone name: CH230-77J24  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 225556 bases at least Q40  
Consensus quality: 228083 bases at least Q30  
Consensus quality: 229291 bases at least Q20  
Estimated insert size: 229928; sum-of-contigs estimation  
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
consists of 2 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
be preserved.

\* 1 236030: contig of 236030 bp in length  
\* 236031 236130: gap of unknown length  
\* 236131 237447: contig of 1317 bp in length.  
Location/Qualifiers  
1. .237447  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"

/db\_xref="taxon:10116"  
/clone="CH230-77J24"  
1. .1266  
/note="wgs\_contig"  
4757. .6020  
/note="wgs\_contig"  
69963. .71541  
/note="wgs\_contig"  
236031. .236130  
/estimated\_length=unknown

ORIGIN  
Query Match 91.6%; Score 17.4; DB 14; Length 237447;  
Best Local Similarity 94.7%; Pred. No. 1.6e+02; Indels 0; Gaps 0;  
Matches 18; Conservative 0; Mismatches 1;  
QY 1 CCACGTGCATCCAAATTGA 19  
|||||  
Db 202315 CCACGTGCTCCAAATTGA 202297  
|||||

RESULT 22  
AC132525/c  
LOCUS AC132525 237900 bp DNA linear HTG 19-NOV-2002  
DEFINITION Rattus norvegicus clone CH230-96B9, WORKING DRAFT SEQUENCE, 2  
unordered pieces.  
AC132525  
AC132525.3 GI:25073428  
HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
KEYWORDS Rattus norvegicus (Norway rat)  
SOURCE  
ORGANISM  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE  
1 (bases 1 to 237900)  
AUTHORS  
Muzny, D., Marle, H., Metzker, M., Lee, S., Amin, A., Angiano, D.,  
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Angiano, D.,  
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,  
Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Z., Chu, J.,  
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,  
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
Gebregorgis, B., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,  
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,  
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,  
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Liu, J.,  
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
Lorensuhera, L., Loulseghe, H., Lozado, R.J., Lu, X., Ma, J.,  
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,  
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,  
Milosavljevic, A., Miner, G., Minja, E., Montanayor, J., Moore, S.,  
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,  
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
Nwackemele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,  
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,  
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.,  
Puaio, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,  
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,  
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,

FEATURES  
source



Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## Direct Submission

## Unpublished

## 2 (bases 1 to 237900)

## Rat Genome Sequencing Consortium.

## Direct Submission

Submitted (01-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

## 3 (bases 1 to 237900)

## Rat Genome Sequencing Consortium.

## Direct Submission

Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 19, 2002 this sequence version replaced gi:23265611.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

## ----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GURJ

Center clone name: CH230-96E8

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 232020 bases at least Q40

Consensus quality: 233552 bases at least Q30

Consensus quality: 234516 bases at least Q20

Estimated insert size: 236203; sum-of-contigs estimation

Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length

\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 2 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 211651: contig of 211651 bp in length

\* 211652 211751: gap of unknown length

\* 211752 237900: contig of 26149 bp in length.

Location/Qualifiers

1..237900

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clone="CH230-96E8"

## FEATURES

## source

misc\_feature 1..1016  
/note="wgs end extension  
clone end:Sp6"  
1244..2473  
/note="wgs end extension  
clone end:Sp6"  
6432..7315  
/note="clone boundary  
clone end:Sp6  
site:\_\_\_\_\_  
end sequence:BH273691"  
211652..211751  
/estimated\_length=unknown

gap

ORIGIN

Query Match 91.6%; Score 17.4; DB 14; Length 237900;  
Best Local Similarity 94.7%; Pred. No. 1.6e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATTGA 19  
|||||  
Db 56811 CCCAGTGCCTCCAAATTGA 56793  
|||||

RESULT 23  
AC122628  
LOCUS  
DEFINITION  
Rattus norvegicus clone CH230-101F19, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, 3 unordered pieces.  
AC122628  
ACCESSION  
VERSION AC122628.4 GI:30579779  
KEYWORDS  
HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
SOURCE  
Rattus norvegicus (Norway rat)  
ORGANISM  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Muridae; Murinae; Rattus.  
REFERENCE  
1 (bases 1 to 244647)  
Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyie, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Frazer, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gregorogis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Haylak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowals, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Louisegh, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mathew, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidas, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nair, L., Nwaokemele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,

Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umani,K., Valae,R., Vera,V., Villaeana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,X., Zhou,S., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

Direct Submission  
Unpublished  
2 (bases 1 to 244647)  
Worley,K.C.

Direct Submission  
Submitted (25-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 244647)

Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On May 13, 2003 this sequence version replaced gi:23270349.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GKCL  
Center clone name: CH230-101F19  
----- Summary Statistics  
Assembly program: Atlas 3.0;  
Consensus quality: 225584 bases at least Q40  
Consensus quality: 228215 bases at least Q30  
Consensus quality: 229693 bases at least Q20  
Estimated insert size: 241268; sum-of-contigs estimation  
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 3 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved  
\* 1 240638: contig of 240638 bp in length  
\* 240639 242657: contig of 1919 bp in length  
\* 242658 242757: gap of unknown length  
\* 242758 244647: contig of 1890 bp in length.  
\* Location/Qualifiers

source  
1..244647  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-101F19"  
1..2330  
/note="wgs end extension  
clone end:Sp6"  
6993..7873  
/note="clone boundary  
clone end:Sp6  
site:EcoRI  
end sequence:BH310527"  
151043..153860  
/note="wgs contig"  
236564..237450  
/note="clone boundary  
clone end:T7  
site:EcoRI  
end sequence:BH310525"  
238071..240638  
/note="wgs end extension  
clone end:T7"  
240639..240738  
/estimated\_length=unknown  
242658..242757  
/estimated\_length=unknown

misc\_feature  
misc\_feature  
misc\_feature  
misc\_feature

gap  
gap

ORIGIN  
Query Match 91.6%; Score 17.4; DB 14; Length 244647;  
Best Local Similarity 94.7%; Pred. No. 1.6e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CCCAGTGCATCCAAATTGA 19  
|||||  
Db 68840 CCCAGTGAATCCAAATTGA 68858

RESULT 24  
AC097593/c  
LOCUS  
DEFINITION Rattus norvegicus clone CH230-162J2, WORKING DRAFT SEQUENCE, 10  
AC097593 251288 bp DNA linear HTG 10-MAY-2003  
unordered pieces.  
AC097593  
AC097593.7 GI:30522795  
HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
KEYWORDS Rattus norvegicus (Norway rat)  
SOURCE Rattus norvegicus  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridea; Muridae; Murinae; Rattus.  
REFERENCE 1 (bases 1 to 251288)  
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Ayodeji,M., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,O., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Frazer,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Geunardis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guervara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,



Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,  
 Kowis,C., Kraft,C.B., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,  
 Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,  
 Lorensuewa,L., Louisseg,H., Lozado,R.J., Lu,X., Ma,J.,  
 Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,  
 Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,  
 Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,  
 Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,  
 Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,  
 Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,  
 Nwaokeleneh,O., Okwuonu,G., Olarunpungoon,A., Pal,S., Parks,K.,  
 Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,  
 Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.,  
 Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,  
 Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,  
 Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,  
 Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,  
 Shetty,J., Shvartabeyn,A., Sisson,I., Sitter,C.D., Smajls,D.,  
 Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,  
 Steinfle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,  
 Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,  
 Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,  
 Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,  
 Williams,G., Willson,R., Wlarczyk,R., Wooden,H., Worley,K.,  
 Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,  
 Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von  
 Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,  
 Weinstock,G. and Gibbs,R.A.

## TITLE

Direct Submission

## JOURNAL

Unpublished

## REFERENCE

2 (bases 1 to 251288)

## AUTHORS

Worley,K.C.

## TITLE

Direct Submission

## JOURNAL

Submitted (20-OCT-2001)

Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

## REFERENCE

3 (bases 1 to 251288)

## AUTHORS

Rat Genome Sequencing Consortium.

## TITLE

Direct Submission

## JOURNAL

Submitted (10-MAY-2003)

Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

## COMMENT

On May 10, 2003 this sequence version replaced gi:33269163.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
 ----- Project Information  
 Center project name: GGJT  
 Center clone name: CH230-162J2  
 ----- Summary Statistics  
 Assembly program: Atlas 3.0;  
 Consensus quality: 205739 bases at least Q40  
 Consensus quality: 209081 bases at least Q30  
 Consensus quality: 211454 bases at least Q20  
 Estimated insert size: 225296; sum-of-contigs estimation  
 Quality coverage: 7x in Q20 bases; sum-of-contigs estimation  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 10 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 4399: contig of 4399 bp in length  
 4400 4499: gap of unknown length  
 4500 233372: contig of 228873 bp in length  
 233373 233472: gap of unknown length  
 233473 235287: contig of 1815 bp in length  
 235288 235387: gap of unknown length  
 235388 236443: contig of 1056 bp in length  
 236444 236543: gap of unknown length  
 236544 239640: contig of 3097 bp in length  
 239641 239740: gap of unknown length  
 239741 241399: contig of 1559 bp in length  
 241400 243026: contig of unknown length  
 243027 243126: gap of unknown length  
 243127 244589: contig of 1464 bp in length  
 244590 244689: gap of unknown length  
 244690 247751: contig of 3061 bp in length  
 247752 251288: gap of unknown length  
 251289 251388: contig of 3438 bp in length.

## FEATURES

## Source

1..251288

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clone="CH230-162J2"

1..1697

/note="wgs contig"

4400..4499

/estimated\_length=unknown

233373..233472

/estimated\_length=unknown

235288..235387

/estimated\_length=unknown

236444..236543

/estimated\_length=unknown

239641..239740

/estimated\_length=unknown

241300..241399

/estimated\_length=unknown

243026..243125

/estimated\_length=unknown

244590..244689

/estimated\_length=unknown

247751..247850

/estimated\_length=unknown

## ORIGIN

Query Match 91.6%; Score 17.4; DB 14; Length 251288;

Best Local Similarity 94.7%; Pred. No. 1.6e+02;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCAGTGCATCCAAATTGA 19

| | | | | | | | | | | | | | | | | | | | |

Db 191493 CGCAGTGCATCCAAATTGA 191475

## RESULT 25

AC116265/c

LOCUS

DEFINITION

AC116265

AC116265

AC116265.3

VERSION

KEYWORDS

SOURCE

AC116265 276127 bp DNA linear HTG 22-SEP-2002  
 Rattus norvegicus clone CH230-18307, \*\*\* SEQUENCING IN PROGRESS  
 \*\*\* 4 unordered pieces.  
 AC116265  
 AC116265.3 GI:23194527  
 HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
 Rattus norvegicus (Norway rat)

## ORGANISM

Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Muridae; Murinae; Rattus.

REFERENCE  
AUTHORS

1 (bases 1 to 276127)  
Muzny, D. Marie., Metzker, M. Lee., Abranzon, S., Adams, C., Alder, J.,  
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,  
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,  
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
Egan, A., Escoto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,  
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
Gebregregios, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,  
Harvey, Y., Haylak, P., Hawes, A., Henderson, N., Hernandez, J.,  
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,  
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,  
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
Kowals, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
Liu, J., Liu, Y., Liu, Y., London, P., Longacre, S., Lopez, J.,  
Lorensuhead, L., Loulsegod, H., Lozada, R. J., Lu, X., Ma, J.,  
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,  
Mangun, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
Mawhiney, S., McLeod, M. P., McNeill, T. Z., Meenen, B.,  
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,  
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
Nwaokeme, O., Okwuonu, G., Olarnpungoon, A., Pal, S., Parks, K.,  
Pasternak, S., Paul, H., Perez, A., Perez, L., Prannkoch, C.,  
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L. L.,  
Puafo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,  
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
Rivas, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,  
Sanders, W., Saverly, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,  
Shetty, J., Shwartzbeyn, A., Sisson, I., Sitter, C. D., Smajs, D.,  
Sneed, A., Sodergren, E., Song, X. Z., Sorelle, R., Sosa, J.,  
Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,  
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,  
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,  
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
Williams, G., Willson, R., Wléczyk, R., Wooden, H., Worley, K.,  
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
Yu, F., Zhang, J., Zhou, J., Zhou, J., Zhou, S., Zhao, S., Dunn, D., von  
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,  
Weinstock, G. and Gibbs, R. A.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Direct Submission  
Unpublished  
2 (bases 1 to 276127)  
Worley, K. C.  
Direct Submission  
Submitted (26-MAR-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 276127)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (22-SEP-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Sep 19, 2002 this sequence version replaced gi:21745788.  
The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the  
sequence may extend beyond the ends of the clone and there may be

contigs that consist entirely of whole genome shotgun sequence  
reads. Both end sequences and whole genome shotgun sequence only  
contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
-----  
Center project name: GFRS  
Center clone name: CH230-18307  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 193244 bases at least Q40  
Consensus quality: 198270 bases at least Q30  
Consensus quality: 201828 bases at least Q20  
Estimated insert size: 216402; sum-of-contigs estimation  
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation  
-----

\* NOTE: Estimated insert size may differ from sequence length  
(see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).  
\* NOTE: This is a 'working draft' sequence. It currently  
consists of 4 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

\* 1 61990: contig of 61990 bp in length  
\* 61991 62090: gap of unknown length  
\* 62091 272212: contig of 210122 bp in length  
\* 272213 272312: gap of unknown length  
\* 272313 273570: contig of 1258 bp in length  
\* 273571 273670: gap of unknown length  
\* 273671 276127: contig of 2457 bp in length.

FEATURES  
source

1..276127  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-18307"

## misc\_feature

56582..56720  
/note="clone boundary  
clone\_end:Sp6  
site:ECORI"

## misc\_feature

end sequence: BH345007"  
complement(60939..61801)  
/note="clone\_boundary  
clone\_end:17  
site:ECORI"

## gap

end sequence: BH345006"  
61991..62090

## gap

/estimated\_length=unknown  
272213..272312

## gap

/estimated\_length=unknown  
273571..273670  
/estimated\_length=unknown

## ORIGIN

Query Match 91.6%; Score 17.4; DB 14; Length 276127;  
Best Local Similarity 94.7%; Pred. No. 1.6e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 CCCAGTGCATCCAAATTGA 19  
|||||  
Db 119572 CCCAGTGCATCCAAATTGA 119554

## COMMENT

RESULT 26  
AC123403  
LOCUS AC123403 Rattus norvegicus clone CH230-291G17, WORKING DRAFT SEQUENCE, 2  
DEFINITION HTG 19-NOV-2002

unordered pieces.  
AC123403  
VERSION  
HTG; HTGS PHASE1; HTGS DRAFT; HTGS\_FULLTOP.  
SOURCE  
Rattus norvegicus (Norway rat)  
ORGANISM  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Rattus.  
REFERENCE  
1 (bases 1 to 278063)  
Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,  
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Angiano, D.,  
Anyalebech, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,  
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,  
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
Frazer, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,  
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogue, M.,  
Hollins, B., Howells, S., Hulyk, S., Hume, J., Iglebird, D., Jackson, A.,  
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
Kowls, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
Lorensuewa, L., Loulsegged, H., Lozado, R. J., Lu, X., Ma, J.,  
Mahenwari, M., Maindarte, M., Mahmoud, M., Malloy, K., Mangum, A.,  
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, B.,  
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
Morgan, M., Morris, K., Morris, S., Muidasa, M., Murphy, M., Nair, L.,  
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K.,  
Nwabelemeh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,  
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,  
Plopper, F., Polindexter, A., Popovic, D., Primus, E., Pu, L.,  
Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,  
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,  
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,  
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D.,  
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,  
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,  
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,  
Valas, R., Vera, V., Villasaana, D., Waldron, L., Walker, B., Wang, J.,  
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
Williams, G., Willson, R., Wlaczek, R., Woodson, H., Worley, K.,  
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,  
Weinstock, G. and Gibbs, R. A.  
Direct Submission  
Unpublished  
2 (bases 1 to 278063)  
Worley, K. C.  
Direct Submission  
Submitted (29-MAY-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 278063)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

On Nov 19, 2002 this sequence version replaced gi:22856217.  
The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described  
in the feature table below represents a scaffold in the Atlas  
assembly (a 'contig-scaffold'). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented, and separated  
by sized gaps filled with Ns to the estimated size. The sequence  
may extend beyond the ends of the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole  
genome shotgun sequence reads. Both end sequences and whole genome  
shotgun sequence only contigs will be indicated in the feature  
table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GXAS  
Center clone name: CH230-291G17  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 262709 bases at least Q40  
Consensus quality: 265699 bases at least Q30  
Consensus quality: 267413 bases at least Q20  
Estimated insert size: 265506; sum-of-contigs estimation  
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html)  
\* NOTE: This sequence may represent more than one clone.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 2 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 276871: contig of 276871 bp in length  
\* 276872 276971: gap of unknown length  
\* 276972 278063: contig of 1092 bp in length.

FEATURES  
source  
1. .278063  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-291G17"  
misc\_feature  
1. .1418  
/note="wgs contig"  
misc\_feature  
4304..5898  
/note="wgs contig"  
misc\_feature  
5979..7130  
/note="wgs contig"  
misc\_feature  
273297..274567  
/note="wgs contig"  
misc\_feature  
275526..276871  
/note="wgs contig"  
gap  
276872..276971  
/estimated\_length=unknown

ORIGIN  
Query Match 91.6%; Score 17.4; DB 14; Length 278063;  
Best Local Similarity 94.7%; Pred. No. 1.6e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATTGA 19  
|||||  
Db 258624 CCCAGTGCATCCAAATTGA 258642  
|||||

RESULT 27

```

AC146012
LOCUS       AC146012             314016 bp    DNA        linear    HTG 01-AUG-2003
DEFINITION Pan troglodytes chromosome UNK clone RP43-124117, *** SEQUENCING IN
PROGRESS ***, 64 unordered pieces.
AC146012
VERSION     AC146012.1  GI:33386964
KEYWORDS    HTG; HTGS_PHASE1
SOURCE      Pan troglodytes (chimpanzee)
ORGANISM    Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homidae; Pan.
            1 (bases 1 to 314016)
REFERENCE   Wilson.R.K.
AUTHORS     The sequence of Pan troglodytes clone
TITLE       Unpublished
JOURNAL
REFERENCE   2 (bases 1 to 314016)
AUTHORS     Wilson.R.K.
TITLE       Direct Submission
JOURNAL     Submitted (01-AUG-2003) Genetics, Genome Sequencing Center, 4444
            Forest Park Parkway, St. Louis, MO 63108, USA

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: C PT124117
----- Summary Statistics -----
Sequencing vector: M13; 0%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 291041 bases at least Q40
Consensus quality: 295402 bases at least Q30
Consensus quality: 298383 bases at least Q20
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 64 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 3952: contig of 3952 bp in length
* 4052: gap of unknown length
* 4053
* 7651: contig of 3599 bp in length
* 7652
* 7752
* 11939: contig of 4188 bp in length
* 11940
* 12039: gap of unknown length
* 12040
* 16200: contig of 4161 bp in length
* 16201
* 16300: gap of unknown length
* 16301
* 22325: contig of 6025 bp in length
* 22326
* 22425: gap of unknown length
* 22426
* 25693: contig of 3268 bp in length
* 25694
* 39155: contig of 13362 bp in length
* 39156
* 39255: gap of unknown length
* 39256
* 69439: contig of 30084 bp in length
* 69440
* 124450: gap of unknown length
* 124451
* 124550: contig of 55011 bp in length
* 124551
* 210270: contig of 85720 bp in length
* 210271
* 210370: gap of unknown length
* 210371
* 212144: contig of 1774 bp in length
* 212244: gap of unknown length
* 213521: contig of 1277 bp in length
* 213522
* 213621: gap of unknown length
* 213622
* 214667: contig of 1046 bp in length
* 214668
* 214767: gap of unknown length

214768
* 216007: contig of 1240 bp in length
* 216008
* 216107: gap of unknown length
* 217221: contig of 1114 bp in length
* 217222
* 218555: contig of 1534 bp in length
* 218556
* 218955: gap of unknown length
* 218956
* 220360: contig of 1405 bp in length
* 220361
* 220460: gap of unknown length
* 220461
* 221975: contig of 1515 bp in length
* 221976
* 222075: gap of unknown length
* 222076
* 223444: contig of 1369 bp in length
* 223445
* 223544: gap of unknown length
* 223545
* 225108: contig of 1564 bp in length
* 225109
* 225208: gap of unknown length
* 225209
* 227073: contig of 1865 bp in length
* 227074
* 227173: gap of unknown length
* 227174
* 228727: contig of 1554 bp in length
* 228728
* 228827: gap of unknown length
* 228828
* 229988: contig of 1161 bp in length
* 229989
* 230088: gap of unknown length
* 230089
* 231297: contig of 1209 bp in length
* 231298
* 231397: gap of unknown length
* 231398
* 232866: contig of 1469 bp in length
* 232867
* 232966: gap of unknown length
* 232967
* 234502: contig of 1536 bp in length
* 234503
* 234603: gap of unknown length
* 234604
* 236778: contig of 2176 bp in length
* 236779
* 236878: gap of unknown length
* 236879
* 238049: contig of 1171 bp in length
* 238050
* 238149: gap of unknown length
* 238150
* 239745: contig of 1596 bp in length
* 239746
* 239845: gap of unknown length
* 239846
* 241223: contig of 1378 bp in length
* 241224
* 241323: gap of unknown length
* 241324
* 242649: contig of 1326 bp in length
* 242650
* 242749: gap of unknown length
* 242750
* 244061: contig of 1312 bp in length
* 244062
* 244161: gap of unknown length
* 244162
* 245558: contig of 1797 bp in length
* 245559
* 246058: gap of unknown length
* 246059
* 247272: contig of 1214 bp in length
* 247273
* 247372: gap of unknown length
* 247373
* 248741: contig of 1369 bp in length
* 248742
* 248841: gap of unknown length
* 248842
* 250480: contig of 1639 bp in length
* 250481
* 250580: gap of unknown length
* 250581
* 252069: contig of 1489 bp in length
* 252070
* 252169: gap of unknown length
* 252170
* 254869: contig of 2700 bp in length
* 254870
* 254969: gap of unknown length
* 254970
* 256315: contig of 1946 bp in length
* 256316
* 257015: gap of unknown length
* 257016
* 257499: contig of 1734 bp in length
* 258750
* 258849: gap of unknown length
* 258850
* 260874: contig of 2025 bp in length
* 260875
* 260876
* 260877: gap of unknown length
* 260878
* 262044: contig of 1070 bp in length
* 262045
* 262144: gap of unknown length
* 262145
* 263605: contig of 1461 bp in length
* 263606
* 263705: gap of unknown length
* 263706
* 263707: contig of 1654 bp in length
* 263708
* 265459: gap of unknown length
* 265460
* 265461
* 267318: contig of 2459 bp in length
* 267319
* 268018: gap of unknown length
* 268019
* 270579: contig of 2561 bp in length
* 270580
* 270679: gap of unknown length
* 270680
* 272170: contig of 1491 bp in length
* 272171
* 272270: gap of unknown length
* 272271
* 273800: contig of 1530 bp in length
* 273801
* 273900: gap of unknown length
* 273901
* 276289: contig of 2389 bp in length
* 276290
* 276389: gap of unknown length
* 276390
* 278517: contig of 2128 bp in length

```

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

REFERENCE  
AUTHORS

1 Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H., Sekine, M., Oobayashi, M., Nishi, T., Shibahara, T., Tanaka, T., Iehii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y., Nagahari, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M., Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T., Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H., Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M., Hiraoka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S., Hosoda, M., Houta, T., Kusano, J., Kanehori, K., Takahashi-Fujii, A., Hara, H., Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R., Takeuchi, K., Arita, M., Imose, N., Musashino, K., Yuuki, H., Oshima, A., Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T., Shiohata, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S., Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H., Coto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T., Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K., Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y., Ozaki, K., Hirao, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N., Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T., Noguchi, S., Itoh, T., Shigetani, K., Senba, T., Matsumura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T., Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J., Satoh, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K., Nagase, T., Nomura, N., Kikuchi, H., Masuho, Y., Yamashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S.  
Complete sequencing and characterization of 21,243 full-length human cDNAs  
Nat. Genet. 36 (1), 40-45 (2004)  
14702039

## TITLE

## JOURNAL

## PUBMED

REFERENCE  
AUTHORS

2 Oshima, A., Takahashi-Fujii, A., Tanase, T., Imose, N., Takeuchi, K., Arita, M., Musashino, K., Yuuki, H., Hara, H., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahari, K., Masuho, Y., Nagai, K. and Isogai, T.  
NEDO human cDNA sequencing project  
Unpublished  
3 (bases 1 to 3242)  
Isogai, T. and Yamamoto, J.  
Direct Submission  
Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan  
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)  
NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB.  
Location/Qualifiers  
1. .3242  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="TESTI2038958"  
/tissue\_type="testis"  
/clone\_lib="TESTI2"  
/note="cloning vector: pME18SFL3"

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Query Match 91.6%; Score 17.4; DB 14; Length 314016;  
Best Local Similarity 94.7%; Pred. No. 1.6e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CCCAGTGCATCCCAATTGA 19  
|||||  
Db 256855 CCCAGTGCATCCCAATTGA 256873  
RESULT 28  
AK097742/c  
LOCUS AK097742 3242 bp mRNA linear PRI 30-JAN-2004  
DEFINITION Homo sapiens cDNA FLJ40423 fis, clone TESTI2038958.  
ACCESSION AK097742  
VERSION AK097742.1 GI:21757605  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens (human)

FEATURES  
source

1. .314016  
Location/Qualifiers  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
/chromosome="UNK"  
/clone="RP43-124117"  
1. .3952  
/notes="assembly\_name:Contig100"  
3953. .4052  
/estimated\_length=unknown  
4053. .7651  
/note="assembly\_name:Contig101"  
7652. .7751  
/estimated\_length=unknown  
7752. .11939  
/notes="assembly\_name:Contig102"  
11940. .12039  
/estimated\_length=unknown  
12040. .16200  
/notes="assembly\_name:Contig103"  
16201. .16300  
/estimated\_length=unknown  
16301. .22325  
/notes="assembly\_name:Contig104"  
22326. .22425  
/estimated\_length=unknown

RESULT 28  
AK097742/c

LOCUS AK097742 3242 bp mRNA linear PRI 30-JAN-2004  
DEFINITION Homo sapiens cDNA FLJ40423 fis, clone TESTI2038958.  
ACCESSION AK097742  
VERSION AK097742.1 GI:21757605  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens (human)

## ORIGIN

Query Match 89.5%; Score 17; DB 8; Length 3242;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCcAGTGCATCCAAATT 17  
|||||  
Db 1271 CCcAGTGCATCCAAATT 1255

RESULT 29  
AL137802/c

LOCUS  
DEFINITION Human DNA sequence from clone RP4-798A10 on chromosome 1 Contains the 3', end of a novel gene (FLJ10420), three novel genes, the 5' end of a novel gene and four CpG islands, complete sequence.

ACCESSION AL137802.7 GI:8452475

VERSION AL137802.7

KEYWORDS HTG; FLJ10420.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 88176)  
Howden, P.

AUTHORS Direct Submission

TITLE Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk

JOURNAL Clone requests; clonerequest@sanger.ac.uk

COMMENT On Jun 10, 2000 this sequence version replaced gi:7960570.  
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:  
Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at  
http://www.sanger.ac.uk/Projects/C/elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr1  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: vegas@sanger.ac.uk  
-----

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.  
RP4-798A10 is from the library RPci-4 constructed by the group of Pieter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pCVPAC2.

FEATURES  
source  
1. .88176  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="RZPD:RPCIP704A10798"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="RP4-798A10"  
/clone\_lib="RPci-4"  
1  
misc\_feature  
/note="Clone\_left end: RP4-798A10"  
join(AL669962.4:2901..3016,AL669962.4:5795..5895,  
3315..3419,3505..3586,4538..4646,7283..7460,11263..11338,  
14287..15523)  
/genes="RP4-798A10.1"  
/locus\_tag="RP4-798A10.1-001"

mrna  
join(AL669962.4:2901..3016,AL669962.4:5795..5895,  
3315..3419,3505..3586,4538..4646,7283..7460,11263..11338,  
14287..15523)  
/gene="RP4-798A10.1"  
/locus\_tag="RP4-798A10.1-001"  
/product="novel protein"  
/note="match: CDNAs: Em:BC018914.1"  
join(AL669962.4:2921..3016,AL669962.4:5795..5895,  
3315..3419,3505..3586,4538..7460,11263..11338,  
14287..15522)  
/gene="RP4-798A10.1"  
/locus\_tag="RP4-798A10.1-003"  
join(AL669962.4:2906..3016,AL669962.4:5795..5895,  
3315..3419,3505..3586,4538..4646,7283..7460,11263..11338,  
14287..14507,14719..15522)  
/gene="RP4-798A10.1"  
/locus\_tag="RP4-798A10.1-002"  
join(AL669962.4:2921..3016,AL669962.4:5795..5895,  
3315..3419,3505..3586,4538..7460,11263..11338,  
14287..15522)  
/gene="RP4-798A10.1"  
/locus\_tag="RP4-798A10.1-003"  
/product="novel protein"  
/note="match: CDNAs: Em:AK024468.1"  
join(AL669962.4:2906..3016,AL669962.4:5795..5895,  
3315..3419,3505..3586,4538..4646,7283..7460,11263..11338,  
14287..14507,14719..15522)  
/gene="RP4-798A10.1"  
/locus\_tag="RP4-798A10.1-002"  
/product="novel protein"  
/note="match: CDNAs: Em:AK001282.1"  
join(AL669962.4:2919..3016,AL669962.4:5795..5895,  
3315..3419,3505..3586,4538..4646,5570..5686,7283..7460,  
11263..11338,14287..14401)  
/gene="RP4-798A10.1"  
/locus\_tag="RP4-798A10.1-005"  
join(AL669962.4:2919..3016,AL669962.4:5795..5895,  
3315..3419,3505..3586,4538..4646,5570..5686,7283..7460,  
11263..11338,14287..14401)  
/gene="RP4-798A10.1-005"  
/locus\_tag="RP4-798A10.1-005"  
/product="novel protein"  
/note="match: ESTs: Em:BX377358.1"  
join(AL669962.4:2925..3016,AL669962.4:5795..5895,  
3315..3419,3505..3586,4538..4646,7283..7460,11263..11338,  
14287..14335)  
/gene="RP4-798A10.1"  
/locus\_tag="RP4-798A10.1-002"  
/standard\_name="OTTHUMP00000002442"  
/codon\_start=1  
/product="novel protein"  
/protein\_id="CAI2843.1"  
/db\_xref="GI:56204279"  
/db\_xref="UniProt/Swiss-Prot:Q9NV23"  
/translation="MEESGYESVLCVKPDVHYRIPPRATNRGYRAAEWQLDQPSWSG  
RLRITAKGMAYIKLEDRTSGBLFAQPDVDPFGTAVESVTDSSRVFVRIEDGNRR  
AFIGIGFGRGDAFDNFVALQDHFVKVQCCFEAKOQNPDOGPKLDLGFKEGQTIKL  
NIANWKKKEGAGNRPVPASTGGSLLPFPFGKGTSTLIPPGEQLA VGGSLVQPAV  
APSSGGA PVPWPQNPATDIWDFTKSTSSQTQPGTGWQF"  
join(AL669962.4:2925..3016,AL669962.4:5795..5895,  
3315..3419,3505..3586,4538..4646,5570..5665)  
/gene="RP4-798A10.1"  
/locus\_tag="RP4-798A10.1-005"  
/standard\_name="OTTHUMP00000002444"  
/codon\_start=1  
/product="novel protein"  
/protein\_id="CAI2842.1"  
/db\_xref="GI:56204278"  
/db\_xref="UniProt/Swiss-Prot:Q9NV23"  
/translation="MEESGYESVLCVKPDVHYRIPPRATNRGYRAAEWQLDQPSWSG  
RLRITAKGMAYIKLEDRTSGBLFAQPDVDPFGTAVESVTDSSRVFVRIEDGNRR  
AFIGIGFGRGDAFDNFVALQDHFVKVQCCFEAKOQNPDOGPKLDLGFKEGQTIKL  
NIANWKKKEGAGNRPVPASTGGSLLPFPFGKGTSTLIPPGEQLA VGGSLVQPAV  
APSSGGA PVPWPQNPATDIWDFTKSTSSQTQPGTGWQF"  
join(AL669962.4:2925..3016,AL669962.4:5795..5895,  
3315..3419,3505..3586,4538..4646,5570..5665)  
/gene="RP4-798A10.1"  
/locus\_tag="RP4-798A10.1-005"  
/standard\_name="OTTHUMP00000002444"  
/codon\_start=1  
/product="novel protein"  
/protein\_id="CAI2843.1"  
/db\_xref="GI:56204278"  
/db\_xref="UniProt/Swiss-Prot:Q9NV23"  
/translation="MEESGYESVLCVKPDVHYRIPPRATNRGYRAAEWQLDQPSWSG  
RLRITAKGMAYIKLEDRTSGBLFAQPDVDPFGTAVESVTDSSRVFVRIEDGNRR  
AFIGIGFGRGDAFDNFVALQDHFVKVQCCFEAKOQNPDOGPKLDLGFKEGQTIKL  
NIANWKKKEGAGNRPVPASTGGSLLPFPFGKGTSTLIPPGEQLA VGGSLVQPAV  
APSSGGA PVPWPQNPATDIWDFTKSTSSQTQPGTGWQF"  
join(AL669962.4:2925..3016,AL669962.4:5795..5895,  
3315..3419,3505..3586,4538..4646,5570..5665)  
/gene="RP4-798A10.1"  
/locus\_tag="RP4-798A10.1-005"  
/standard\_name="OTTHUMP00000002444"  
/codon\_start=1  
/product="novel protein"  
/protein\_id="CAI2842.1"  
/db\_xref="GI:56204278"  
/db\_xref="UniProt/Swiss-Prot:Q9NV23"  
/translation="MEESGYESVLCVKPDVHYRIPPRATNRGYRAAEWQLDQPSWSG  
RLRITAKGMAYIKLEDRTSGBLFAQPDVDPFGTAVESVTDSSRVFVRIEDGNRR  
AFIGIGFGRGDAFDNFVALQDHFVKVQCCFEAKOQNPDOGPKLDLGFKEGQTIKL  
NIANWKKKEGAGNRPVPASTGGSLLPFPFGKGTSTLIPPGEQLA VGGSLVQPAV  
APSSGGA PVPWPQNPATDIWDFTKSTSSQTQPGTGWQF"  
join(AL669962.4:2925..3016,AL669962.4:5795..5895,  
3315..3419,3505..3586,4538..4646,5570..5665)  
/gene="RP4-798A10.1"  
/locus\_tag="RP4-798A10.1-005"  
/standard\_name="OTTHUMP00000002444"  
/codon\_start=1  
/product="novel protein"  
/protein\_id="CAI2842.1"  
/db\_xref="GI:56204278"  
/db\_xref="UniProt/Swiss-Prot:Q9NV23"  
/translation="MEESGYESVLCVKPDVHYRIPPRATNRGYRAAEWQLDQPSWSG  
RLRITAKGMAYIKLEDRTSGBLFAQPDVDPFGTAVESVTDSSRVFVRIEDGNRR  
AFIGIGFGRGDAFDNFVALQDHFVKVQCCFEAKOQNPDOGPKLDLGFKEGQTIKL  
NIANWKKKEGAGNRPVPASTGGSLLPFPFGKGTSTLIPPGEQLA VGGSLVQPAV  
APSSGGA PVPWPQNPATDIWDFTKSTSSQTQPGTGWQF"  
join(AL669962.4:2925..3016,AL669962.4:5795..5895,  
3315..3419,3505..3586,4538..4646,5570..5665)  
/gene="RP4-798A10.1"  
/locus\_tag="RP4-798A10.1-005"  
/standard\_name="OTTHUMP00000002444"  
/codon\_start=1  
/product="novel protein"  
/protein\_id="CAI2842.1"  
/db\_xref="GI:56204278"  
/db\_xref="UniProt/Swiss-Prot:Q9NV23"  
/translation="MEESGYESVLCVKPDVHYRIPPRATNRGYRAAEWQLDQPSWSG  
RLRITAKGMAYIKLEDRTSGBLFAQPDVDPFGTAVESVTDSSRVFVRIEDGNRR  
AFIGIGFGRGDAFDNFVALQDHFVKVQCCFEAKOQNPDOGPKLDLGFKEGQTIKL  
NIANWKKKEGAGNRPVPASTGGSLLPFPFGKGTSTLIPPGEQLA VGGSLVQPAV  
APSSGGA PVPWPQNPATDIWDFTKSTSSQTQPGTGWQF"  
join(AL669962.4:2925..3016,AL669962.4:5795..5895,  
3315..3419,3505..3586,4538..4646,5570..5665)  
/gene="RP4-798A10.1"  
/locus\_tag="RP4-798A10.1-005"  
/standard\_name="OTTHUMP00000002444"  
/codon\_start=1  
/product="novel protein"  
/protein\_id="CAI2842.1"  
/db\_xref="GI:56204278"  
/db\_xref="UniProt/Swiss-Prot:Q9NV23"  
/translation="MEESGYESVLCVKPDVHYRIPPRATNRGYRAAEWQLDQPSWSG  
RLRITAKGMAYIKLEDRTSGBLFAQPDVDPFGTAVESVTDSSRVFVRIEDGNRR  
AFIGIGFGRGDAFDNFVALQDHFVKVQCCFEAKOQNPDOGPKLDLGFKEGQTIKL  
NIANWKKKEGAGNRPVPASTGGSLLPFPFGKGTSTLIPPGEQLA VGGSLVQPAV  
APSSGGA PVPWPQNPATDIWDFTKSTSSQTQPGTGWQF"  
join(AL669962.4:2925..3016,AL669962.4:5795..5895,  
3315..3419,3505..3586,4538..4646,5570..5665)  
/gene="RP4-798A10.1"  
/locus\_tag="RP4-798A10.1-005"  
/standard\_name="OTTHUMP00000002444"  
/codon\_start=1  
/product="novel protein"  
/protein\_id="CAI2842.1"  
/db\_xref="GI:56204278"  
/db\_xref="UniProt/Swiss-Prot:Q9NV23"  
/translation="MEESGYESVLCVKPDVHYRIPPRATNRGYRAAEWQLDQPSWSG  
RLRITAKGMAYIKLEDRTSGBLFAQPDVDPFGTAVESVTDSSRVFVRIEDGNRR  
AFIGIGFGRGDAFDNFVALQDHFVKVQCCFEAKOQNPDOGPKLDLGFKEGQTIKL  
NIANWKKKEGAGNRPVPASTGGSLLPFPFGKGTSTLIPPGEQLA VGGSLVQPAV  
APSSGGA PVPWPQNPATDIWDFTKSTSSQTQPGTGWQF"  
join(AL669962.4:2925..3016,AL669962.4:5795..5895,  
3315..3419,3505..3586,4538..4646,5570..5665)  
/gene="RP4-798A10.1"  
/locus\_tag="RP4-798A10.1-005"  
/standard\_name="OTTHUMP00000002444"  
/codon\_start=1  
/product="novel protein"  
/protein\_id="CAI2842.1"  
/db\_xref="GI:56204278"  
/db\_xref="UniProt/Swiss-Prot:Q9NV23"  
/translation="MEESGYESVLCVKPDVHYRIPPRATNRGYRAAEWQLDQPSWSG  
RLRITAKGMAYIKLEDRTSGBLFAQPDVDPFGTAVESVTDSSRVFVRIEDGNRR  
AFIGIGFGRGDAFDNFVALQDHFVKVQCCFEAKOQNPDOGPKLDLGFKEGQTIKL  
NIANWKKKEGAGNRPVPASTGGSLLPFPFGKGTSTLIPPGEQLA VGGSLVQPAV  
APSSGGA PVPWPQNPATDIWDFTKSTSSQTQPGTGWQF"  
join(AL669962.4:2925..3016,AL669962.4:5795..5895,  
3315..3419,3505..3586,4538..4646,5570..5665)  
/gene="RP4-798A10.1"  
/locus\_tag="RP4-798A10.1-005"  
/standard\_name="OTTHUMP00000002444"  
/codon\_start=1  
/product="novel protein"  
/protein\_id="CAI2842.1"  
/db\_xref="GI:56204278"  
/db\_xref="UniProt/Swiss-Prot:Q9NV23"  
/translation="MEESGYESVLCVKPDVHYRIPPRATNRGYRAAEWQLDQPSWSG  
RLRITAKGMAYIKLEDRTSGBLFAQPDVDPFGTAVESVTDSSRVFVRIEDGNRR  
AFIGIGFGRGDAFDNFVALQDHFVKVQCCFEAKOQNPDOGPKLDLGFKEGQTIKL  
NIANWKKKEGAGNRPVPASTGGSLLPFPFGKGTSTLIPPGEQLA VGGSLVQPAV  
APSSGGA PVPWPQNPATDIWDFTKSTSSQTQPGTGWQF"  
join(AL669962.4:2925..3016,AL669962.4:5795..5895,  
3315..3419,3505..3586,4538..4646,5570..5665)  
/gene="RP4-798A10.1"  
/locus\_tag="RP4-798A10.1-005"  
/standard\_name="OTTHUMP00000002444"  
/codon\_start=1  
/product="novel protein"  
/protein\_id="CAI2842.1"  
/db\_xref="GI:56204278"  
/db\_xref="UniProt/Swiss-Prot:Q9NV23"  
/translation="MEESGYESVLCVKPDVHYRIPPRATNRGYRAAEWQLDQPSWSG  
RLRITAKGMAYIKLEDRTSGBLFAQPDVDPFGTAVESVTDSSRVFVRIEDGNRR  
AFIGIGFGRGDAFDNFVALQDHFVKVQCCFEAKOQNPDOGPKLDLGFKEGQTIKL  
NIANWKKKEGAGNRPVPASTGGSLLPFPFGKGTSTLIPPGEQLA VGGSLVQPAV  
APSSGGA PVPWPQNPATDIWDFTKSTSSQTQPGTGWQF"  
join(AL669962.4:2925..3016,AL669962.4:5795..5895,  
3315..3419,3505..3586,4538..4646,5570..5665)  
/gene="RP4-798A10.1"  
/locus\_tag="RP4-798A10.1-005"  
/standard\_name="OTTHUMP00000002444"  
/codon\_start=1  
/product="novel protein"  
/protein\_id="CAI2842.1"  
/db\_xref="GI:56204278"  
/db\_xref="UniProt/Swiss-Prot:Q9NV23"  
/translation="MEESGYESVLCVKPDVHYRIPPRATNRGYRAAEWQLDQPSWSG  
RLRITAKGMAYIKLEDRTSGBLFAQPDVDPFGTAVESVTDSSRVFVRIEDGNRR  
AFIGIGFGRGDAFDNFVALQDHFVKVQCCFEAKOQNPDOGPKLDLGFKEGQTIKL  
NIANWKKKEGAGNRPVPASTGGSLLPFPFGKGTSTLIPPGEQLA VGGSLVQPAV  
APSSGGA PVPWPQNPATDIWDFTKSTSSQTQPGTGWQF"  
join(AL669962.4:2925..3016,AL669962.4:5795..5895,  
3315..3419,3505..3586,4538..4646,5570..5665)  
/gene="RP4-798A10.1"  
/locus\_tag="RP4-798A10.1-005"  
/standard\_name="OTTHUMP00000002444"  
/codon\_start=1  
/product="novel protein"  
/protein\_id="CAI2842.1"  
/db\_xref="GI:56204278"  
/db\_xref="UniProt/Swiss-Prot:Q9NV23"  
/translation="MEESGYESVLCVKPDVHYRIPPRATNRGYRAAEWQLDQPSWSG  
RLRITAKGMAYIKLEDRTSGBLFAQPDVDPFGTAVESVTDSSRVFVRIEDGNRR  
AFIGIGFGRGDAFDNFVALQDHFVKVQCCFEAKOQNPDOGPKLDLGFKEGQTIKL  
NIANWKKKEGAGNRPVPASTGGSLLPFPFGKGTSTLIPPGEQLA VGGSLVQPAV  
APSSGGA PVPWPQNPATDIWDFTKSTSSQTQPGTGWQF"  
join(AL669962.4:2925..3016,AL669962.4:5795..5895,  
3315..3419,3505..3586,4538..4646,5570..5665)  
/gene="RP4-798A10.1"  
/locus\_tag="RP4-798A10.1-005"  
/standard\_name="OTTHUMP00000002444"  
/codon\_start=1  
/product="novel protein"  
/protein\_id="CAI2842.1"  
/db\_xref="GI:56204278"  
/db\_xref="UniProt/Swiss-Prot:Q9NV23"  
/translation="MEESGYESVLCVKPDVHYRIPPRATNRGYRAAEWQLDQPSWSG  
RLRITAKGMAYIKLEDRTSGBLFAQPDVDPFGTAVESVTDSSRVFVRIEDGNRR  
AFIGIGFGRGDAFDNFVALQDHFVKVQCCFEAKOQNPDOGPKLDLGFKEGQTIKL  
NIANWKKKEGAGNRPVPASTGGSLLPFPFGKGTSTLIPPGEQLA VGGSLVQPAV  
APSSGGA PVPWPQNPATDIWDFTKSTSSQTQPGTGWQF"  
join(AL669962.4:2925..3016,AL669962.4:5795..5895,  
3315..3419,3505..3586,4538..4646,5570..5665)  
/gene="RP4-798A10.1"  
/locus\_tag="RP4-798A10.1-005"  
/standard\_name="OTTHUMP00000002444"  
/codon\_start=1  
/product="novel protein"  
/protein\_id="CAI2842.1"  
/db\_xref="GI:56204278"  
/db\_xref="UniProt/Swiss-Prot:Q9NV23"  
/translation="MEESGYESVLCVKPDVHYRIPPRATNRGYRAAEWQLDQPSWSG  
RLRITAKGMAYIKLEDRTSGBLFAQPDVDPFGTAVESVTDSSRVFVRIEDGNRR  
AFIGIGFGRGDAFDNFVALQDHFVKVQCCFEAKOQNPDOGPKLDLGFKEGQTIKL  
NIANWKKKEGAGNRPVPASTGGSLLPFPFGKGTSTLIPPGEQLA VGGSLVQPAV  
APSSGGA PVPWPQNPATDIWDFTKSTSSQTQPGTGWQF"  
join(AL669962.4:2925..3016,AL669962.4:5795..5895,  
3315..3419,3505..3586,4538..4646,5570..5665)  
/gene="RP4-798A10.1"  
/locus\_tag="RP4-798A10.1-005"  
/standard\_name="OTTHUMP00000002444"  
/codon\_start=1  
/product="novel protein"  
/protein\_id="CAI2842.1"  
/db\_xref="GI:56204278"  
/db\_xref="UniProt/Swiss-Prot:Q9NV23"  
/translation="MEESGYESVLCVKPDVHYRIPPRATNRGYRAAEWQLDQPSWSG  
RLRITAKGMAYIKLEDRTSGBLFAQPDVDPFGTAVESVTDSSRVFVRIEDGNRR  
AFIGIGFGRGDAFDNFVALQDHFVKVQCCFEAKOQNPDOGPKLDLGFKEGQTIKL  
NIANWKKKEGAGNRPVPASTGGSLLPFPFGKGTSTLIPPGEQLA VGGSLVQPAV  
APSSGGA PVPWPQNPATDIWDFTKSTSSQTQPGTGWQF"  
join(AL669962.4:2925..3016,AL669962.4:5795..5895,  
3315..3419,3505..3586,4538..4646,5570..5665)  
/gene="RP4-798A10.1"  
/locus\_tag="RP4-798A10.1-005"  
/standard\_name="OTTHUMP00000002444"  
/codon\_start=1  
/product="novel protein"  
/protein\_id="CAI2842.1"  
/db\_xref="GI:56204278"  
/db\_xref="UniProt/Swiss-Prot:Q9NV23"  
/translation="MEESGYESVLCVKPDVHYRIPPRATNRGYRAAEWQLDQPSWSG  
RLRITAKGMAYIKLEDRTSGBLFAQPDVDPFGTAVESVTDSSRVFVRIEDGNRR  
AFIGIGFGRGDAFDNFVALQDHFVKVQCCFEAKOQNPDOGPKLDLGFKEGQTIKL  
NIANWKKKEGAGNRPVPASTGGSLLPFPFGKGTSTLIPPGEQLA VGGSLVQPAV  
APSSGGA PVPWPQNPATDIWDFTKSTSSQTQPGTGWQF"  
join(AL669962.4:2925..3016,AL669962.4:5795..5895,  
3315..3419,3505..3586,4538..4646,5570..5665)  
/gene="RP4-798A10.1"  
/locus\_tag="RP4-798A10.1-005"  
/standard\_name="OTTHUMP00000002444"  
/codon\_start=1  
/product="novel protein"  
/protein\_id="CAI2842.1"  
/db\_xref="GI:56204278"  
/db\_xref="UniProt/Swiss-Prot:Q9NV23"  
/translation="MEESGYESVLCVKPDVHYRIPPRATNRGYRAAEWQLDQPSWSG  
RLRITAKGMAYIKLEDRTSGBLFAQPDVDPFGTAVESVTDSSRVFVRIEDGNRR  
AFIGIGFGRGDAFDNFVALQDHFVKVQCCFEAKOQNPDOGPKLDLGFKEGQTIKL  
NIANWKKKEGAGNRPVPASTGGSLLPFPFGKGTSTLIPPGEQLA VGGSLVQPAV  
APSSGGA PVPWPQNPATDIWDFTKSTSSQTQPGTGWQF"  
join(AL669962.4:2925..3016,AL669962.4:5795..5895,  
3315..3419,3505..3586,4538..4646,5570..5665)  
/gene="RP4-798A10.1"  
/locus\_tag="RP4-798A10.1-005"  
/standard\_name="OTTHUMP00000002444"  
/codon\_start=1  
/product="novel protein"  
/protein\_id="CAI2842.1"  
/db\_xref="GI:56204278"  
/db\_xref="UniProt/Swiss-Prot:Q9NV23"  
/translation="MEESGYESVLCVKPDVHYRIPPRATNRGYRAAEWQLDQPSWSG  
RLRITAKGMAYIKLEDRTSGBLFAQPDVDPFGTAVESVTDSSRVFVRIEDGNRR  
AFIGIGFGRGDAFDNFVALQDHFVKVQCCFEAKOQNPDOGPKLDLGFKEGQTIKL  
NIANWKKKEGAGNRPVPASTGGSLLPFPFGKGTSTLIPPGEQLA VGGSLVQPAV  
APSSGGA PVPWPQNPATDIWDFTKSTSSQTQPGTGWQF"  
join(AL669962.4:2925..3016,AL669962.4:5795..5895,  
3315..3419,3505..3586,4538..4646,5570..5665)  
/gene="RP4-798A10.1"  
/locus\_tag="RP4-798A10.1-005"  
/standard\_name="OTTHUMP00000002444"  
/codon\_start=1  
/product="novel protein"  
/protein\_id="CAI2842.1"  
/db\_xref="GI:56204278"  
/db\_xref="UniProt/Swiss-Prot:Q9NV23"  
/translation="MEESGYESVLCVKPDVHYRIPPRATNRGYRAAEWQLDQPSWSG  
RLRITAKGMAYIKLEDRTSGBLFAQPDVDPFGTAVESVTDSSRVFVRIEDGNRR  
AFIGIGFGRGDAFDNFVALQDHFVKVQCCFEAKOQNPDOGPKLDLGFKEGQTIKL  
NIANWKKKEGAGNRPVPASTGGSLLPFPFGKGTSTLIPPGEQLA VGGSLVQPAV  
APSSGGA PVPWPQNPATDIWDFTKSTSSQTQPGTGWQF"  
join(AL669962.4:2925..3016,AL669962.4:5795..5895,  
3315..3419,3505..3586,4538..4646,5570..5665)  
/gene="RP4-798A10.1"  
/locus\_tag="RP4-798A10.1-005"  
/standard\_name="OTTHUMP00000002444"  
/codon\_start=1  
/product="novel protein"  
/protein\_id="CAI2842.1"  
/db\_xref="GI:56204278"  
/db\_xref="UniProt/Swiss-Prot:Q9NV23"  
/translation="MEESGYESVLCVKPDVHYRIPPRATNRGYRAAEWQLDQPSWSG  
RLRITAKGMAYIKLEDRTSGBLFAQPDVDPFGTAVESVTDSSRVFVRIEDGNRR  
AFIGIGFGRGDAFDNFVALQDHFVKVQCCFEAKOQNPDOGPKLDLGFKEGQTIKL  
NIANWKKKEGAGNRPVPASTGGSLLPFPFGKGTSTLIPPGEQLA VGGSLVQPAV  
APSSGGA PVPWPQNPATDIWDFTKSTSSQTQPGTGWQF"  
join(AL669962.4:2925..3016,AL669962.4:5795..5895,  
3315..3419,3505..3586,4538..4646,5570..5665)  
/gene="RP4-798A10.1"  
/locus\_tag="RP4-798A10.1-005"  
/standard\_name="OTTHUMP00000002444"  
/codon\_start=1  
/product="novel protein"  
/protein\_id="CAI2842.1"  
/db\_xref="GI:56204278"  
/db\_xref="UniProt/Swiss-Prot:Q9NV23"  
/translation="MEESGYESVLCVKPDVHYRIPPRATNRGYRAAEWQLDQPSWSG  
RLRITAKGMAYIKLEDRTSGBLFAQPDVDPFGTAVESVTDSSRVFVRIEDGNRR  
AFIGIGFGRGDAFDNFVALQDHFVKVQCCFEAKOQNPDOGPKLDLGFKEGQTIKL  
NIANWKKKEGAGNRPVPASTGGSLLPFPFGKGTSTLIPPGEQLA VGGSLVQPAV  
APSSGGA PVPWPQNPATDIWDFTKSTSSQTQPGTGWQF"  
join(AL669962.4:2925..3016,AL669962.4:5795..5895,  
3315..3419,3505..3586,4538..4646,5570..5665)  
/gene="RP4-798A10.1"  
/locus\_tag="RP4-798A10.1-005"  
/standard\_name="OTTHUMP00000002444"  
/codon\_start=1  
/product="novel protein"  
/protein\_id="CAI2842.1"  
/db\_xref="GI:56204278"  
/db\_xref="UniProt/Swiss-Prot:Q9NV23"  
/translation="MEESGYESVLCVKPDVHYRIPPRATNRGYRAAEWQLDQPSWSG  
RLRITAKGMAYIKLEDRTSGBLFAQPDVDPFGTAVESVTDSSRVFVRIEDGNRR  
AFIGIGFGRGDAFDNFVALQDHFVKVQCCFEAKOQNPDOGPKLDLGFKEGQTIKL  
NIANWKKKEGAGNRPVPASTGGSLLPFPFGKGTSTLIPPGEQLA VGGSLVQPAV  
APSSGGA PVPWPQNPATDIWDFTKSTSSQTQPGTGWQF"  
join(AL669962.4:2925..3016,AL669962.4:5795..5895,  
3315..3419,3505..3586,4538..4646,5570..5665)  
/gene="RP4-798A10.1"  
/locus\_tag="RP4-798A10.1-005"  
/standard\_name="OTTHUMP00000002444"  
/codon\_start=1  
/product="novel protein"  
/protein\_id="CAI2842.1"  
/db\_xref="GI:56204278"  
/db\_xref="UniProt/Swiss-Prot:Q9NV23"  
/translation="MEESGYESVLCVKPDVHYRIPPRATNRGYRAAEWQLDQPSWSG  
RLRITAKGMAYIKLEDRTSGBLFAQPDVDPFGTAVESVTDSSRVFVRIEDGNRR  
AFIGIGFGRGDAFDNFVALQDHFVKVQCCFEAKOQNPDOGPKLDLGFKEGQTIKL  
NIANWKKKEGAGNRPVPASTGGSLLPFPFGKGTSTLIPPGEQLA VGGSLVQPAV  
APSSGGA PVPWPQNPATDIWDFTKSTSSQTQPGTGWQF"  
join(AL669962.4:2925..3016,AL669962.4:5795..5895,  
3315..3419,3505..3586,4538..4646,5570..5665)  
/gene="RP4-798A10.1"  
/locus\_tag="RP4-798A10.1-005"  
/standard\_name="OTTHUMP00000002444"  
/codon\_start=1  
/product="novel protein"  
/protein\_id="CAI2842.1"  
/db\_xref="GI:56204278"  
/db\_xref="UniProt/Swiss-Prot:Q9NV23"  
/translation="MEESGYESVLCVKPDVHYRIPPRATNRGYRAAEWQLDQPSWSG  
RLRITAKGMAYIKLEDRTSGBLFAQPDVDPFGTAVESVTDSSRVFVRIEDGNRR  
AFIGIGFGRGDAFDNFVALQDHFVKVQCCFEAKOQNPDOGPKLDLGFKEGQTIKL  
NIANWKKKEGAGNRPVPASTGGSLLPFPFGKGTSTLIPPGEQLA VGGSLVQPAV  
APSSGGA PVPWPQNPATDIWDFTKSTSSQTQPGTGWQF"  
join(AL669962.4:2925..3016,AL669962.4:5795..5895,  
3315..3419,3505..3586,4538..4646,5570..5665)  
/gene="RP4-798A10.1"  
/locus\_tag="RP4-798A10.1-005"  
/standard\_name="OTTHUMP00000002444"  
/codon\_start=1  
/product="novel protein"  
/protein\_id="CAI2842.1"  
/db\_xref="GI:56204278"  
/db\_xref="UniProt/Swiss-Prot:Q9NV23"  
/translation="MEESGYESVLCVKPDVHYRIPPRATNRGYRAAEWQLDQPSWSG  
RLRITAKGMAYIKLEDRTSGBLFAQPDVDPFGTAVESVTDSSRVFVRIEDGNRR  
AFIGIGFGRGDAFDNFVALQDHFVKVQCCFEAKOQNPDOGPKLDLGFKEGQTIKL  
NIANWKKKEGAGNRPVPASTGGSLLPFPFGKGTSTLIPPGEQLA VGGSLVQPAV  
APSSGGA PVPWPQNPATDIWDFTKSTSSQTQPGTGWQF"  
join(AL669962.4:2925..3016,AL669962.4:5795..5895,  
3315..3419,3505..3586,4538..4646,5570..5665)  
/gene="RP4-798A10.1"  
/locus\_tag





```

/note="assembly_fragment:06035
clone_end:SP6
vector_side:right
clone_end:T7
vector_side:left"
/misc_feature
/9354..269489
/note="assembly_fragment:07223"
/misc_feature
/9354..269489
/note="assembly_fragment:07668"
/misc_feature
/269590..377192
/note="assembly_fragment:07692"
ORIGIN

```

```

Query Match      89.5%; Score 17; DB 14; Length 110000;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 CCAGTGCATCCAAATT 17
          |||||
Db      1178 CCAGTGCATCCAAATT 1162

```

```

RESULT 31
AC015618/c
LOCUS      AC015618      144605 bp      DNA      linear      HTG 26-MAY-2000
DEFINITION Homo sapiens clone RP11-4513, WORKING DRAFT SEQUENCE, 20 unordered
            pieces.
ACCESSION      AC015618
VERSION      AC015618.3 GI:8096819
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens

```

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.

```

```

REFERENCE
AUTHORS      Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE      Homo sapiens chromosome, clone RP11-4513
JOURNAL      Unpublished

```

```

REFERENCE
AUTHORS      2 (bases 1 to 144605)
            Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavskiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArelano,K., Dewar,K., Domino,M., Donean,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heatford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,K., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 26, 2000 this sequence version replaced gi:6563632.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1159
Center clone name: 45_1_3
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731

```

# FEATURES

```

source
1..144605
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="RP11-4513"
/clone_lib="RP11-11 Human Male BAC"
1..322
/note="assembly_fragment
clone end:T7
vector side:right"
323..422
/estimated_length=100
423..1731
/note="assembly_fragment"
1732..1831
/estimated_length=100
1832..3125

misc_feature
1..322
misc_feature
323..422
gap
323..422
misc_feature
423..1731
gap
1732..1831
misc_feature
1832..3125

```

```

Consensus quality: 118070 bases at least Q40
Consensus quality: 132878 bases at least Q30
Consensus quality: 138494 bases at least Q20
Insert size: 157000; agarose-fp
Quality coverage: 3.5 in Q20 bases; agarose-fp
Quality coverage: 3.9 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

1
322: contig of 322 bp in length
323
422: gap of 100 bp
423
1731: contig of 1309 bp in length
1732
1831: gap of 100 bp
1832
1325: contig of 1294 bp in length
3126
3225: gap of 100 bp
3226
5623: contig of 2398 bp in length
5624
5723: gap of 100 bp
5724
9305: contig of 3582 bp in length
9306
9405: gap of 100 bp
9406
14253: contig of 4848 bp in length
14254
14353: gap of 100 bp
14354
16991: contig of 2638 bp in length
16992
17091: gap of 100 bp
17092
22044: contig of 4953 bp in length
22045
22144: gap of 100 bp
22145
26351: contig of 4207 bp in length
26352
26451: gap of 100 bp
26452
31059: contig of 4608 bp in length
31060
31159: gap of 100 bp
31160
37358: contig of 6199 bp in length
37359
37458: gap of 100 bp
37459
43606: contig of 6148 bp in length
43607
43707
49619: contig of 5913 bp in length
49620
49719: gap of 100 bp
49720
57388: contig of 7669 bp in length
57389
57488: gap of 100 bp
57489
63726: contig of 6238 bp in length
63727
63826: gap of 100 bp
63827
70666: contig of 6840 bp in length
70667
70766: gap of 100 bp
70767
78969: contig of 8203 bp in length
78970
79069: gap of 100 bp
79070
96162: contig of 17093 bp in length
96163
96262: gap of 100 bp
96263
117346: contig of 21084 bp in length
117347
117446: gap of 100 bp
117447
144605: contig of 27159 bp in length.

```

## FEATURES

```

source
1..144605
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="RP11-4513"
/clone_lib="RP11-11 Human Male BAC"
1..322
/note="assembly_fragment
clone end:T7
vector side:right"
323..422
/estimated_length=100
423..1731
/note="assembly_fragment"
1732..1831
/estimated_length=100
1832..3125

misc_feature
1..322
misc_feature
323..422
gap
323..422
misc_feature
423..1731
gap
1732..1831
misc_feature
1832..3125

```



gap /note="assembly\_fragment"  
3126. .3225  
/estimated\_length=100  
misc\_feature 3226. .5623  
/note="assembly\_fragment"  
gap 5624. .5723  
/estimated\_length=100  
misc\_feature 5724. .9305  
/note="assembly\_fragment"  
gap 9306. .9405  
/estimated\_length=100  
misc\_feature 9406. .14253  
/note="assembly\_fragment"  
gap 14254. .14353  
/estimated\_length=100  
misc\_feature 14354. .16591  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:left  
gap 16992. .17091  
/estimated\_length=100  
misc\_feature 17092. .22044  
/note="assembly\_fragment"  
gap 22045. .22144  
/estimated\_length=100  
misc\_feature 22145. .26351  
/note="assembly\_fragment"  
gap 26352. .26451  
/estimated\_length=100  
misc\_feature 26452. .31059  
/note="assembly\_fragment"  
gap 31060. .31159  
/estimated\_length=100  
misc\_feature 31160. .37358  
/note="assembly\_fragment"  
gap 37359. .37458  
/estimated\_length=100  
misc\_feature 37459. .43606  
/note="assembly\_fragment"  
gap 43607. .43706  
/estimated\_length=100  
misc\_feature 43707. .49619  
/note="assembly\_fragment"  
gap 49620. .49719  
/estimated\_length=100  
misc\_feature 49720. .57388  
/note="assembly\_fragment"  
gap 57389. .57488  
/estimated\_length=100  
misc\_feature 57489. .63726  
/note="assembly\_fragment"  
gap 63727. .63826  
/estimated\_length=100  
misc\_feature 63827. .70666  
/note="assembly\_fragment"  
gap 70667. .70766  
/estimated\_length=100  
misc\_feature 70767. .78969  
/note="assembly\_fragment"  
gap 78970. .79069  
/estimated\_length=100  
misc\_feature 79070. .96162  
/note="assembly\_fragment"  
gap 96163. .96262  
/estimated\_length=100  
misc\_feature 96263. .117346  
/note="assembly\_fragment"  
gap 117347. .117446  
/estimated\_length=100  
misc\_feature 117447. .144605  
/note="assembly\_fragment"

ORIGIN

Query Match 89.5%; Score 17; DB 14; Length 144605;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 CCCAGTGCATCCAAATT 17  
Db 15486 CCCAGTGCATCCAAATT 15470  
  
RESULT 32  
AL691480  
LOCUS AL691480  
DEFINITION Mus musculus chromosome X clone RP23-11608, WORKING DRAFT SEQUENCE,  
14 unordered pieces.  
ACCESSION AL691480  
VERSION AL691480.2 GI:19699754  
KEYWORDS HTG; HTGS PHASE1; HTGS\_DRAFT.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridea; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 170869)  
McLay, K.  
Direct Submission  
Submitted (22-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Mar 24, 2002 this sequence version replaced gi:19572674.  
COMMENT  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquery@sanger.ac.uk  
----- Project Information  
Center project name: bm11608  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Consensus quality: 167320 bases at least Q40  
Consensus quality: 168431 bases at least Q30  
Consensus quality: 168994 bases at least Q20  
Insert size: 169569; sum-of-contigs  
Quality coverage: 7.55x in Q20 bases; sum-of-contigs Quality  
coverage: 7.59x in Q20 bases; agarose-fp  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 14 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 22703: contig of 22703 bp in length  
\* 22704 22803: gap of 100 bp  
\* 22804 30610: contig of 7807 bp in length  
\* 30611 30710: gap of 100 bp  
\* 30711 95842: contig of 65132 bp in length  
\* 95843 95942: gap of 100 bp  
\* 95943 100837: contig of 4895 bp in length  
\* 100838 100937: gap of 100 bp  
\* 100938 105345: contig of 4408 bp in length  
\* 105346 105445: gap of 100 bp  
\* 105446 108517: contig of 3072 bp in length  
\* 108518 108617: gap of 100 bp  
\* 108618 119578: contig of 10961 bp in length  
\* 119579 119678: gap of 100 bp  
\* 119679 123510: contig of 3832 bp in length  
\* 123511 123610: gap of 100 bp  
\* 123611 129936: contig of 6326 bp in length  
\* 129937 130036: gap of 100 bp

\* 130037 132928: contig of 2892 bp in length  
 \* 132929 133028: gap of 100 bp  
 \* 133029 154299: contig of 21271 bp in length  
 \* 154300 154399: gap of 100 bp  
 \* 154400 157324: contig of 2925 bp in length  
 \* 157325 157424: gap of 100 bp  
 \* 157425 168178: contig of 10754 bp in length  
 \* 157426 168278: gap of 100 bp  
 \* 168179 170869: contig of 2591 bp in length.

## FEATURES

source  
 1. .170869  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="X"  
 /clone="RP23-11608"  
 /clone\_lib="RPCI-23"

misc\_feature  
 1. .22703  
 /notes="assembly fragment:01006  
 fragment\_chain:1"  
 misc\_feature  
 22804. .30610  
 /notes="assembly fragment:00997  
 fragment\_chain:1"  
 misc\_feature  
 30711. .95842  
 /notes="assembly fragment:02491  
 fragment\_chain:1"  
 misc\_feature  
 95943. .100837  
 /notes="assembly fragment:00409  
 fragment\_chain:1"  
 misc\_feature  
 100938. .105345  
 /notes="assembly fragment:02169  
 fragment\_chain:1"  
 misc\_feature  
 105446. .108517  
 /notes="assembly fragment:01337  
 fragment\_chain:2"  
 misc\_feature  
 108618. .119578  
 /notes="assembly fragment:01405  
 fragment\_chain:2"  
 misc\_feature  
 119679. .123510  
 /notes="assembly fragment:01497  
 fragment\_chain:2"  
 misc\_feature  
 123611. .129936  
 /notes="assembly fragment:00895"  
 misc\_feature  
 130037. .132928  
 /notes="assembly fragment:01048"  
 misc\_feature  
 133029. .154299  
 /notes="assembly fragment:02316.0"  
 misc\_feature  
 154400. .157324  
 /notes="assembly fragment:02892"  
 misc\_feature  
 157425. .168178  
 /notes="assembly fragment:02656  
 fragment\_chain:3"  
 misc\_feature  
 168279. .170869  
 /notes="assembly fragment:00887  
 fragment\_chain:3  
 clone\_end:SP6  
 vector\_side:right"

## ORIGIN

Query Match 89.5% Score 17; DB 14; Length 170869;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATT 17  
 |||||

Db 153126 CCCAGTGCATCCAAATT 153142

## RESULT 33

AL355800  
 LOCUS AL355800 203055 bp DNA linear HTG 10-JUL-2001  
 DEFINITION Homo sapiens chromosome 1 clone RP11-284017, 9 unordered pieces.  
 ACCESSION AL355800

VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AL355800.5 GI:9797364  
 HTG; HTGS\_PHASE1; HTGS\_CANCELLED.  
 Homo sapiens (human)

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Hominidae; Homo.

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
 requests: clonerequest@sanger.ac.uk  
 On Aug 12, 2000 this sequence version replaced gi:9213640.

## COMMENT

----- Genome Center  
 Center: Sanger Centre  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: humquery@sanger.ac.uk  
 ----- Project Information  
 Center project name: BA284017  
 ----- Summary Statistics  
 Assembly program: XGAP4; version 4.5  
 Sequencing vector: plasmid; L08752; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Consensus quality: 199739 bases at least Q40  
 Consensus quality: 200469 bases at least Q30  
 Consensus quality: 200876 bases at least Q20  
 Insert size: 202255; sum-of-contigs  
 Insert size: 198980; agarose-fp  
 Quality coverage: 5.86x in Q20 bases; sum-of-contigs Quality  
 coverage: 6.00x in Q20 bases; agarose-fp  
 -----

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 9 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 6940: contig of 6940 bp in length  
 \* 6941 7040: gap of 100 bp  
 \* 7041 54773: contig of 47733 bp in length  
 \* 54774 54873: gap of 100 bp  
 \* 54874 59278: contig of 4405 bp in length  
 \* 59279 59378: gap of 100 bp  
 \* 59379 115220: contig of 55842 bp in length  
 \* 115221 115320: gap of 100 bp  
 \* 115321 129073: contig of 13753 bp in length  
 \* 129074 129173: gap of 100 bp  
 \* 129174 143521: contig of 14348 bp in length  
 \* 143522 143621: gap of 100 bp  
 \* 143622 146068: contig of 2447 bp in length  
 \* 146069 146168: gap of 100 bp  
 \* 146169 152457: contig of 6289 bp in length  
 \* 152458 152557: gap of 100 bp  
 \* 152558 203055: contig of 50498 bp in length.

## FEATURES

## source

1. .203055  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /clone="RP11-284017"  
 /clone\_lib="RPCI-11.1"

## misc\_feature

1. .6940  
 /note="assembly\_fragment:00858  
 clone\_end:SP6  
 vector\_side:left"

## misc\_feature

7041. .54773

## misc\_feature

/note="assembly\_fragment:00299.0"

54874. .59278

```

/note="assembly_fragment:00999"
59379..115220
/note="assembly_fragment:01107"
115321..129073
/note="assembly_fragment:01203"
129174..143521
/note="assembly_fragment:01233
fragment_chain:1"
143622..146068
/note="assembly_fragment:01770
fragment_chain:1"
146169..152457
/note="assembly_fragment:01108
fragment_chain:1"
152558..203055
/note="assembly_fragment:01951
fragment_chain:1
clone_end:T7
vector_side:right"

ORIGIN
Query Match      89.5%; Score 17; DB 14; Length 203055;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATT 17
|||||
Db 39651 CCCAGTGCATCCAAATT 39667

RESULT 34
AC161653/c
LOCUS
DEFINITION
Bos taurus clone CH240-105K14, *** SEQUENCING IN PROGRESS ***, 25
unordered pieces.
AC161653
AC161653.2 GI:68303057
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
Bos taurus
SOURCE
Bos taurus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Bovinae; Bos.
Muzny,D,Marle., Metzker,M,Lees., Abramson,S., Amin,A., Anguiano,D.,
1 (bases 1 to 275252)
Allen,C., Allen,H., Alsbrooks,S., Ayodeji,M., Baca,E., Baden,H.,
Anyalebechi,V., Ayodeji,A., Ayodeji,M., Barber,M., Barnstead,M., Benahmed,F.,
Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Derano,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,I., Garza,M.,
Georegeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,M., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,R., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoquee,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,W., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Loulsegged,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja.E., Montemayor,J., Moore,S.,

```

```

Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.,
Puafo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Rose,M., Rose,R., Ruiz,S.J.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,F., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villaseana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,P.,
Williams,G., Willson,R., Wlaczek,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,E., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.

Unpublished
Direct Submission
2 (bases 1 to 275252)
Worley,K.C.
Direct Submission
Submitted (18-MAY-2005) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 275252)
Cow Genome Sequencing Consortium.
Direct Submission
Submitted (01-JUL-2005) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jun 29, 2005 this sequence version replaced gi:56267807.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: FGON
Center clone name: CH240-105K14
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 263412 bases at least Q40
Consensus quality: 266191 bases at least Q30
Consensus quality: 268357 bases at least Q20
Estimated insert size: 270842; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will

```



LOCUS CS070928 359 bp DNA linear PAT 05-MAY-2005  
DEFINITION Sequence 91 from Patent WO2001032927.  
ACCESSION CS070928  
VERSION CS070928.1 GI:63088332  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
REFERENCE 1  
AUTHORS Sornasse, T., Seilhamer, J.J. and Watson, G.A.  
TITLE Tissue specific genes of diagnostic import  
JOURNAL Patent: WO 2001032927-A 91 10-MAY-2001;  
Incyte Genomics, Inc. (US)  
FEATURES  
Location/Qualifiers  
1..359  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 86.3%; Score 16.4; DB 6; Length 359;  
Best Local Similarity 94.4%; Pred. No. 5.2e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 CCAGTGCATCCAAATTGA 19  
|||||  
Db 220 CCAGTGCATCCAGATTGA 203  
RESULT 37  
AX407370 465 bp DNA linear PAT 14-JUN-2002  
LOCUS AX407370  
DEFINITION Sequence 17 from Patent WO0229103.  
ACCESSION AX407370  
VERSION AX407370.1 GI:21440075  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
REFERENCE 1  
AUTHORS Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G.  
TITLE Gene expression profiles in liver cancer  
JOURNAL Patent: WO 0229103-A 17 11-APR-2002;  
GENE LOGIC INC (US)  
FEATURES  
Location/Qualifiers  
1..465  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
/note="EMBL/GenBank Accession No. AA007395"  
ORIGIN  
Query Match 86.3%; Score 16.4; DB 6; Length 465;  
Best Local Similarity 94.4%; Pred. No. 5.3e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 CCAGTGCATCCAAATTGA 19  
|||||  
Db 398 CCAGTGCATCCAGATTGA 415  
RESULT 38  
CQ612015/c 711 bp DNA linear PAT 03-FEB-2004  
LOCUS CQ612015  
DEFINITION Sequence 39773 from Patent WO0171042.  
ACCESSION CQ612015  
VERSION CQ612015.1 GI:41663323  
KEYWORDS  
SOURCE Drosophila sp.

ORGANISM Drosophila sp.  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1  
AUTHORS Venter, J.C., Adams, M., Li, P.W. and Myers, E.W.  
TITLE Detection kits, such as nucleic acid arrays, for detecting the  
expression of 10,000 or more Drosophila genes and uses thereof  
JOURNAL Patent: WO 0171042-A 39773 27-SEP-2001;  
PE Corporation (NY) (US)  
FEATURES  
Location/Qualifiers  
1..711  
/organism="Drosophila sp."  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:7242"  
ORIGIN  
Query Match 86.3%; Score 16.4; DB 6; Length 711;  
Best Local Similarity 94.4%; Pred. No. 5.3e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 CCAGTGCATCCAAATTGA 19  
|||||  
Db 334 CCAGTGCATCCAAATCGA 317  
RESULT 39  
AR378249/c 1404 bp DNA linear PAT 18-DEC-2003  
LOCUS AR378249  
DEFINITION Sequence 3255 from patent US 6605709.  
ACCESSION AR378249  
VERSION AR378249.1 GI:40081431  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1404)  
AUTHORS Bretton, G.L.  
TITLE Nucleic acid and amino acid sequences relating to Proteus mirabilis  
for diagnostics and therapeutics  
JOURNAL Patent: US 6605709-A 3255 12-AUG-2003;  
Genome Therapeutics Corporation; Waltham, MA  
FEATURES  
Location/Qualifiers  
1..1404  
/organism="unknown"  
/mol\_type="genomic DNA"  
ORIGIN  
Query Match 86.3%; Score 16.4; DB 6; Length 1404;  
Best Local Similarity 94.4%; Pred. No. 5.4e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CCCAGTGCATCCAAATTG 18  
|||||  
Db 824 CCCAGTGCAGCCAAATTG 807  
RESULT 40  
BC022319/c 1557 bp mRNA linear PRI 11-AUG-2005  
LOCUS BC022319  
DEFINITION Homo sapiens alcohol dehydrogenase 4 (class II), pi polypeptide,  
mRNA (cDNA clone MGC:22633 IMAGE:4722568), complete cds.  
ACCESSION BC022319  
VERSION BC022319.1 GI:18490172  
KEYWORDS MGC.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1557)  
AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,

Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Woreley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettner,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.  
Mammalian Gene Collection Program Team  
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
12477932  
2 (bases 1 to 1557)  
NIH MGC Project  
Direct Submission  
Submitted (01-FEB-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Bethesda, MD 20892-2590, USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs@mail.nih.gov](mailto:cgapbs@mail.nih.gov)  
Tissue Procurement: CLONTECH  
cDNA Library Preparation: CLONTECH Laboratories, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www.shgc.stanford.edu>  
Contact: (Dickson, Mark) [mcd@paxil.stanford.edu](mailto:mcd@paxil.stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAL Plate: 37 Row: g Column: 1  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 11496889.

#### FEATURES

Location/Qualifiers  
1..1557  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="MGC:22633 IMAGE:4722568"  
/tissue\_type="Liver"  
/clone\_lib="NIH MGC 76"  
/lab\_host="DH10B"  
/note="Vector: pDNR-LIB"  
1..1557  
/gene="ADH4"  
/note="synonym: ADH-2"  
/db\_xref="GeneID:127"  
/db\_xref="MIM:103740"  
76..1218  
/gene="ADH4"  
/codon\_start=1  
/product="class II alcohol dehydrogenase 4 pi subunit"  
/protein\_id="AAH22319.1"  
/db\_xref="GI:18490173"  
/db\_xref="GeneID:127"  
/db\_xref="MIM:103740"

#### gene

#### CDS

ORIGIN  
Query Match 86.3%; Score 16.4; DB 8; Length 1557;  
Best Local Similarity 94.4%; Pred. No. 5.4e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 CCAGTGCATCCAAATTGA 19  
|||||  
Db 1129 CCAGTGCATCCAGATTGA 1112  
|||||

Search completed: November 20, 2005, 18:34:05  
Job time : 728.494 secs

/translation="MGTKGKVIKCKAAIAWEAGKPLCTIEVEVAPPKAEVRIQIAT  
SICLSDATVIDSGFVLGAVIVGHEAGIVSIGGVTVNVRPGDKVIPLYAPLCRKC  
KFLSPLTNLCGKI SNLKSPASDQOIMEDKTSRCKGPKPVYHFTGTSFTSOYTVVSD  
INAKIDDDANLERVCLGCGFSTGYGAANNKAVTPGSCAVFGLGGVLSAVMGCK  
AAGASRIIGIDINSEKFPVAKALGATDCLNPRDLHKPIQEVILITKGGVDFALDCAG

GSETMKAALDCTTAGWGSCTFIVGAAGSKGLTVFPPELIIGRTINGTFEGGWKSVDSI  
PKLVTDYKKNKFNLDALVTHLPFDKISEAFDLNQGKSIRTILIF"

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2005, 15:03:55 ; Search time 308.691 Seconds  
(without alignments)  
410.213 Million cell updates/sec

-Title: US-10-627-757-27

Perfect score: 19

Sequence: 1 ccagtgcatccaaattga 19

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_21.\*

- 1: Geneseqn1980s.\*
- 2: Geneseqn1990s.\*
- 3: Geneseqn2000s.\*
- 4: Geneseqn2001as.\*
- 5: Geneseqn2001bs.\*
- 6: Geneseqn2002as.\*
- 7: Geneseqn2002bs.\*
- 8: Geneseqn2003as.\*
- 9: Geneseqn2003bs.\*
- 10: Geneseqn2003cs.\*
- 11: Geneseqn2003ds.\*
- 12: Geneseqn2004as.\*
- 13: Geneseqn2004bs.\*
- 14: Geneseqn2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19	100.0	19	ADL14975	ADL14975 Human gla
2	19	100.0	19	ADW86597	ADW86597 PCR prime
3	19	100.0	1116	ADL14956	ADL14956 Human gla
4	19	100.0	1116	ADW86578	ADW86578 Human opt
5	19	100.0	46951	ADL13891	ADL13891 Human opt
6	17	89.5	14484	ABA07270	ABA07270 Human pan
7	17	89.5	14484	AAK99915	AAK99915 Human dig
8	17	89.5	14495	ABA07272	ABA07272 Human pan
9	17	89.5	14495	AAK99917	AAK99917 Human dig
10	16.4	86.3	359	AAH57251	AAH57251 Human liv
11	16.4	86.3	430	ACL54336	ACL54336 Human col
12	16.4	86.3	465	ABN93519	ABN93519 Gene #17
13	16.4	86.3	711	ABL28355	ABL28355 Drosophil
14	16.4	86.3	1404	ADf02970	ADf02970 Bacterial
15	16.4	86.3	1721	ADJ22628	ADJ22628 Human liv
16	16.4	86.3	1973	AAH57459	AAH57459 Human liv
17	16.4	86.3	2711	ABL28354	ABL28354 Drosophil
18	16.4	86.3	2878	ADC02494	ADC02494 DNA encod
19	16.4	86.3	2878	ADf17879	ADf17879 E2 ubiqui

C 20	16.4	86.3	2878	12	ADQ85957	Adq85957 Human tum
C 21	16.4	86.3	2878	13	ACN40857	Acn40857 Tumour-as
C 22	16.4	86.3	2900	10	ADB75604	Adb75604 Prostate
C 23	16.4	86.3	2919	14	ADZ49223	Adz49223 Insulin s
C 24	16.4	86.3	4740	1	AAK91159	Aak91159 pSK111 pr
C 25	16	84.2	445	3	AAA82213	Aaa82213 N. mening
C 26	15.8	83.2	512	4	ABA61066	Abaa61066 Human foe
C 27	15.8	83.2	512	4	AAI40963	Aai40963 Probe #96
C 28	15.8	83.2	512	4	ABA28970	Aba28970 Probe #74
C 29	15.8	83.2	512	4	AAK35248	Aak35248 Human bon
C 30	15.8	83.2	512	4	AAK09359	Aak09359 Human bra
C 31	15.8	83.2	512	4	ABS34992	Abes34992 Human liv
C 32	15.8	83.2	512	6	ABS09663	Abso9663 Human gen
C 33	15.8	83.2	1998	4	AAH15533	Aah15533 Human cdn
C 34	15.8	83.2	2538	14	ADZ36079	Adz36079 Human NFK
C 35	15.8	83.2	2811	14	ADZ36077	Adz36077 Human NFK
C 36	15.8	83.2	2907	9	ADA02614	Ada02614 Human NFK
C 37	15.8	83.2	2907	10	ADB72352	Adb72352 Human NFK
C 38	15.8	83.2	2907	10	ADE82946	Ades82946 Human NFK
C 39	15.8	83.2	2907	10	ADE95862	Ades95862 Human NFK
C 40	15.8	83.2	2907	14	ADV42551	Adv42551 Human psy
C 41	15.8	83.2	3205	14	ADY26351	Ady26351 Human dna
C 42	15.8	83.2	3497	10	ADE25618	Ades25618 Human cdn
C 43	15.8	83.2	3625	3	AAK34912	Aak34912 Human ade
C 44	15.8	83.2	3625	3	AAF21034	Aaf21034 Human low
C 45	15.8	83.2	3625	6	ABZ35010	Abz35010 Human gen

ALIGNMENTS

RESULT 1

ADL14975

ID ADL14975 standard; DNA; 19 BP.

XX AC ADL14975;

XX DT 06-MAY-2004 (first entry)

XX DE Human glaucoma-related optineurin (OPTN) exon 10 PCR primer SF10.

XX KW Human; glaucoma; optineurin; OPTN; diagnosis; PCR; primer; ss.

XX OS Homo sapiens.

XX PN EP1388590-A2.

XX PD 11-FEB-2004.

XX PF 29-JUL-2003; 2003EP-00447201.

XX PR 02-AUG-2002; 2002JP-00226612.

XX PA (SYSM-) SYSNEX CORP.

XX PI Kouchi Y, Masago A, Takahata T;

XX WPI; 2004-146134/15.

XX Gene assay for predicting future onset of glaucoma, particularly primary open angle glaucoma or normal ocular tension glaucoma, comprises detecting a mutation of at least one base of the optineurin gene.

PS Claim 9; SEQ ID NO 27; 31pp; English.

XX The present sequence is that of PCR primer SF10 for exon 10 ADL14956 of the glaucoma-associated gene, OPTN (optineurin) ADL14949. The invention relates to a gene assay method for predicting future onset of primary open angle glaucoma and/or normal ocular tension glaucoma. This involves detecting a mutation in the OPTN gene coding sequence, specifically a substitution of G for A at position 619 and/or a substitution of A for G at position 898 of the OPTN coding sequence. The mutation(s) is detected using a nucleic acid amplification method using primers specific for the

CC different exons of the coding sequence, including primers SF10 and SR10  
 CC ADL14976 for exon 10.

XX SQ Sequence 19 BP; 6 A; 6 C; 3 G; 4 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 19; DB 12; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATTGA 19  
 |||||  
 Db 1 CCCAGTGCATCCAAATTGA 19

RESULT 2  
 ADW86597  
 ID ADW86597 standard; DNA; 19 BP.  
 XX AC  
 AC ADW86597;  
 XX DT 21-APR-2005 (first entry)  
 XX PCR primer used to amplify human optineurin (OPTN) gene exon 10 Seq 27.  
 DE KW glaucoma; optineurin; ophthalmological; ss; ocular disease;  
 KM DNA amplification; genetic marker; PCR; primer.  
 XX OS Homo sapiens.  
 XX JP2005034112-A.  
 PN 10-FEB-2005.  
 PD 29-JUL-2003; 2003JP-00281897.  
 PF 02-AUG-2002; 2002JP-00226612.  
 PR 30-JUN-2003; 2003JP-00188070.  
 XX (TOAI-) TOA IYO DENSHI KK.  
 PA WPI; 2005-156038/17.  
 XX Estimating risk of onset of glaucoma, involves analyzing mutation in any  
 PT one portion of Optic new phosphorous gene, and utilizing analyzed  
 PT mutation as index for estimating risk of onset of glaucoma.

XX Example 1; SEQ ID NO 27; 13pp; Japanese.  
 CC This invention relates to a novel method for estimating the risk of onset  
 CC of glaucoma. Specifically, it refers to a method that involves analyzing  
 CC a mutation in any one portion of an optineurin (OPTN) gene and utilizing  
 CC the analyzed mutation as an index for estimating the risk of onset of  
 CC glaucoma. The present invention describes oligonucleotides to detect  
 CC mutations that hybridize with one or more portions of the OPTN glaucoma  
 CC related gene. Accordingly, it provides oligos that detect A619G and G898A  
 CC mutations in the human OPTN gene of patients at risk of developing  
 CC glaucoma, in particular primary open-angle glaucoma and/or normal tension  
 CC glaucoma. These mutations will not be present in a sample obtained from a  
 CC healthy person who is not at risk of onset of glaucoma. This  
 CC oligonucleotide sequence is a PCR primer used to amplify a human OPTN  
 CC exon of the invention.

XX SQ Sequence 19 BP; 6 A; 6 C; 3 G; 4 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 19; DB 14; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATTGA 19  
 |||||  
 Db 1 CCCAGTGCATCCAAATTGA 19

RESULT 3  
 ADL14956  
 ID ADL14956 standard; DNA; 1116 BP.  
 XX AC ADL14956;  
 XX DT 06-MAY-2004 (first entry)  
 XX Human glaucoma-related optineurin (OPTN) exon 10.  
 DE KW Human; glaucoma; optineurin; OPTN; diagnosis; ds.  
 XX OS Homo sapiens.

XX FH Key Location/Qualifiers  
 FT primer\_bind Complement(251..269)  
 FT /\*tag= a  
 FT /\*note= "Primer SF10"  
 FT exon 501..616  
 FT /\*tag= b  
 FT /\*number= 10  
 FT 765..786  
 FT /\*tag= c  
 FT /\*note= "Primer SR10"

XX EP1388590-A2.  
 XX 11-FEB-2004.  
 XX 29-JUL-2003; 2003EP-00447201.  
 XX 02-AUG-2002; 2002JP-00226612.  
 XX (SYSM-) SYSMEX CORP.

XX Kouchi Y, Masago A, Takahata T;  
 XX WPI; 2004-146134/15.  
 XX Gene assay for predicting future onset of glaucoma, particularly primary  
 PT open angle glaucoma or normal ocular tension glaucoma, comprises  
 PT detecting a mutation of at least one base of the optineurin gene.

XX Example 1; SEQ ID NO 8; 31pp; English.  
 PS The present sequence comprises exon 10 of the glaucoma-associated gene,  
 CC OPTN (optineurin) ADL14949. The invention relates to a gene assay method  
 CC for predicting future onset of primary open angle glaucoma and/or normal  
 CC ocular tension glaucoma. This involves detecting a mutation in the OPTN  
 CC gene coding sequence, specifically a substitution of G for A at position  
 CC 619 and/or a substitution of A for G at position 898 of the OPTN coding  
 CC sequence. The mutation(s) is detected using a nucleic acid amplification  
 CC method using primers specific for the different exons of the coding  
 CC sequence, including primers SF10 ADL14975 and SR10 ADL14976 for exon 10.

XX SQ Sequence 1116 BP; 379 A; 172 C; 230 G; 335 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 19; DB 12; Length 1116;  
 Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATTGA 19  
 |||||  
 Db 251 CCCAGTGCATCCAAATTGA 269

RESULT 4  
 ADW86578  
 ID ADW86578 standard; DNA; 1116 BP.  
 XX AC ADW86578;  
 XX DT 21-APR-2005 (first entry)



```

XX DE Human optineurin (OPTN) gene exon 10 Seq 8.
XX
XX KW glaucoma; optineurin; ophthalmological; ds; ocular disease;
XX KW DNA amplification; genetic marker.
XX
XX OS Homo sapiens.
XX
XX PN JP2005034112-A.
XX
XX PD 10-FEB-2005.
XX
XX PF 29-JUL-2003; 2003JP-00281897.
XX
XX PR 02-AUG-2002; 2002JP-00226612.
XX PR 30-JUN-2003; 2003JP-00188070.
XX
XX PA (TOAI-) TOA IYO DENSHI KK.
XX
XX DR WPI; 2005-156038/17.
XX
XX PT Estimating risk of onset of glaucoma, involves analyzing mutation in any
XX PT one portion of Optic new phosphorous gene, and utilizing analyzed
XX PT mutation as index for estimating risk of onset of glaucoma.
XX
XX PS Example 1; SEQ ID NO 8; 13pp; Japanese.
XX
XX CC This invention relates to a novel method for estimating the risk of onset
XX CC of glaucoma. Specifically, it refers to a method that involves analyzing
XX CC a mutation in any one portion of an optineurin (OPTN) gene and utilizing
XX CC the analyzed mutation as an index for estimating the risk of onset of
XX CC glaucoma. The present invention describes oligonucleotides to detect
XX CC mutations that hybridize with one or more portions of the OPTN glaucoma
XX CC related gene. Accordingly, it provides oligos that detect A619G and G898A
XX CC mutations in the human OPTN gene of patients at risk of developing
XX CC glaucoma, in particular primary open-angle glaucoma and/or normal tension
XX CC glaucoma. These mutations will not be present in a sample obtained from a
XX CC healthy person who is not at risk of onset of glaucoma. This
XX CC polynucleotide sequence is a human OPTN exon of the invention.
XX
XX SQ Sequence 1116 BP; 379 A; 172 C; 230 G; 335 T; 0 U; 0 Other;

Query Match 100.0%; Score 19; DB 14; Length 1116;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCCAAAATTGA 19
Db 251 CCCAGTGCATCCCAAAATTGA 269

RESULT 5
ADE13891
ID ADE13891 standard; DNA; 46951 BP.
XX
XX AC ADE13891;
XX
XX DT 29-JAN-2004 (first entry)
XX
XX DE Human optineurin gene.
XX
XX KW Human; optineurin; ds; gene; ophthalmological;
XX KW single nucleotide polymorphism; SNP; glaucoma;
XX KW progressive ocular hypertensive disorder; glaucoma related disorder.
XX
XX OS Homo sapiens.
XX
XX FH Key Location/Qualifiers
XX FT variation replace(391,G)
XX FT /*tag= a
XX FT /standard_name= "Single nucleotide polymorphism"
XX FT /*tag= b
XX

```

```

FT variation /standard_name= "Single nucleotide polymorphism"
FT /*tag= c
FT replace(709,G)
FT variation /standard_name= "Single nucleotide polymorphism"
FT /*tag= d
FT replace(887,A)
FT variation /standard_name= "Single nucleotide polymorphism"
FT /*tag= e
FT replace(894,T)
FT variation /standard_name= "Single nucleotide polymorphism"
FT /*tag= f
FT replace(1112,C)
FT variation /standard_name= "Single nucleotide polymorphism"
FT /*tag= g
FT replace(987,C)
FT variation /standard_name= "Single nucleotide polymorphism"
FT /*tag= h
FT replace(1505,CC)
FT variation /standard_name= "Single nucleotide polymorphism"
FT /*tag= i
FT replace(1606,A)
FT variation /standard_name= "Single nucleotide polymorphism"
FT /*tag= j
FT replace(2405,T)
FT variation /standard_name= "Single nucleotide polymorphism"
FT /*tag= k
FT replace(2606,G)
FT variation /standard_name= "Single nucleotide polymorphism"
FT /*tag= l
FT replace(3313,A)
FT variation /standard_name= "Single nucleotide polymorphism"
FT /*tag= m
FT replace(3555,TT)
FT variation /standard_name= "Single nucleotide polymorphism"
FT /*tag= n
FT replace(3625,G)
FT variation /standard_name= "Single nucleotide polymorphism"
FT /*tag= o
FT replace(3629,C)
FT variation /standard_name= "Single nucleotide polymorphism"
FT /*tag= p
FT replace(3882,TT)
FT variation /standard_name= "Single nucleotide polymorphism"
FT /*tag= q
FT replace(3988,T)
FT variation /standard_name= "Single nucleotide polymorphism"
FT /*tag= r
FT replace(4452,A)
FT /standard_name= "Single nucleotide polymorphism"
XX
XX US2003190617-A1.
XX
XX PD 09-OCT-2003.
XX
XX PF 06-MAR-2002; 2002US-00091281.
XX
XX PR 06-MAR-2002; 2002US-00091281.
XX
XX PA (SIEE/) SI E.
XX PA (RAYM/) RAYMOND V.
XX PA (MORI/) MORISSETTE J.
XX
XX PI Raymond V, Morissette J, Si E;
XX WPI; 2003-864168/80.
XX
XX PT New nucleic acid sequences of the optineurin gene are useful to detect
XX PT polymorphisms particularly single nucleotide polymorphisms in the
XX PT optineurin promoter to diagnose, prognosis and treat glaucoma and related
XX PT disorders.
XX
XX PS Disclosure; SEQ ID NO 2; 159pp; English.
XX

```

CC The invention relates to an isolated nucleic acid (N1) comprising at  
 CC least 20 but not more than 1500 consecutive nucleotides of the optineurin  
 CC promoter appearing as Adu3890. Also included are the optineurin promoter  
 CC operably linked to a heterologous nucleic acid, a nucleic acid capable of  
 CC detecting a single nucleotide polymorphism (SNP) in the optineurin  
 CC promoter, a host cell comprising the promoter operably linked to a  
 CC heterologous sequence, diagnosing or prognosing glaucoma in a sample  
 CC obtained from a cell or bodily fluid (comprising glaucoma in a sample  
 CC in a promoter region of the optineurin gene, associated with a glaucoma  
 CC phenotype), detecting a SNP sequence variation in a sample containing  
 CC DNA, detecting the presence of an optineurin promoter sequence variation  
 CC in a sample containing DNA, determining the presence or increased  
 CC susceptibility to glaucoma or to a progressive ocular hypertensive  
 CC disorder resulting in loss of visual field in a patient (or the severity  
 CC or progression of glaucoma in a patient, comprising providing  
 CC amplification reaction primers that direct amplification of a selected  
 CC nucleic acid region containing the variation within the optineurin  
 CC promoter and amplifying the DNA) and detecting a polymorphism (comprising  
 CC obtaining a sample containing human genomic DNA, providing a nucleic acid  
 CC capable of detecting a SNP located within an optineurin promoter, and  
 CC detecting the polymorphism). The invention is used to diagnose and  
 CC prognose glaucoma and also to treat glaucoma related disorders. The  
 CC present sequence is the optineurin gene.

XX SQ Sequence 46951 BP; 12703 A; 10108 C; 10051 G; 14089 T; 0 U; 0 Other;

Query Match 100.0%; Score 19; DB 10; Length 46951;

Best Local Similarity 100.0%; Pred. No. 25;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATTGA 19

Db 28589 CCCAGTGCATCCAAATTGA 28607

RESULT 6

ABA07270

ID ABA07270 standard; DNA; 14484 BP.

XX AC ABA07270;

XX DT

14-JAN-2002 (first entry)

XX DE

Human pancreatic cancer related genomic DNA, SEQ ID NO: 589.

XX KW

Human; cytostatic; antidiabetic; antiinflammatory; gastric; osteopathic;

KW anti-hormone; antiulcer; thyroid-active; gene therapy; antisense therapy;

KW pancreatic cancer antigen inhibitor; pancreatic cancer; pancreatitis;

KW diabetes; endocrine disorder; acromegaly; hyperthyroidism;

KW gastrointestinal disorder; Crohn's disease; duodenal ulcer; ds.

XX OS

Homo sapiens.

XX XX

WO200155206-A1.

XX XX

02-AUG-2001.

XX XX

17-JAN-2001; 2001WO-US0001353.

XX XX

31-JAN-2000; 2000US-0179065P.

PR PR

04-FEB-2000; 2000US-0180628P.

PR PR

24-FEB-2000; 2000US-0184664P.

PR PR

02-MAR-2000; 2000US-0186350P.

PR PR

16-MAR-2000; 2000US-0189874P.

PR PR

17-MAR-2000; 2000US-0190076P.

PR PR

18-APR-2000; 2000US-0198123P.

PR PR

19-MAY-2000; 2000US-0205151P.

PR PR

07-JUN-2000; 2000US-0209467P.

PR PR

28-JUN-2000; 2000US-0214886P.

PR PR

11-JUL-2000; 2000US-0217496P.

PR PR

14-JUL-2000; 2000US-0218290P.

PR PR

26-JUL-2000; 2000US-0220963P.

PR PR

26-JUL-2000; 2000US-0220964P.

PR PR

14-AUG-2000; 2000US-0224518P.

PR PR

14-AUG-2000; 2000US-0224519P.

PR PR

14-AUG-2000; 2000US-0225213P.

PR PR

14-AUG-2000; 2000US-0225214P.

PR PR

14-AUG-2000; 2000US-0225266P.

PR PR

14-AUG-2000; 2000US-0225267P.

PR PR

14-AUG-2000; 2000US-0225268P.

PR PR

14-AUG-2000; 2000US-0225270P.

PR PR

14-AUG-2000; 2000US-0225447P.

PR PR

14-AUG-2000; 2000US-0225757P.

PR PR

14-AUG-2000; 2000US-0225758P.

PR PR

14-AUG-2000; 2000US-0225759P.

PR PR

18-AUG-2000; 2000US-0226279P.

PR PR

22-AUG-2000; 2000US-0226681P.

PR PR

22-AUG-2000; 2000US-0226689P.

PR PR

22-AUG-2000; 2000US-0227182P.

PR PR

23-AUG-2000; 2000US-0227009P.

PR PR

30-SEP-2000; 2000US-0228924P.

PR PR

01-SEP-2000; 2000US-0229343P.

PR PR

01-SEP-2000; 2000US-0229344P.

PR PR

01-SEP-2000; 2000US-0229345P.

PR PR

05-SEP-2000; 2000US-0229509P.

PR PR

05-SEP-2000; 2000US-0229513P.

PR PR

06-SEP-2000; 2000US-0230437P.

PR PR

06-SEP-2000; 2000US-0230438P.

PR PR

08-SEP-2000; 2000US-0231242P.

PR PR

08-SEP-2000; 2000US-0231243P.

PR PR

08-SEP-2000; 2000US-0231244P.

PR PR

08-SEP-2000; 2000US-0231413P.

PR PR

08-SEP-2000; 2000US-0231414P.

PR PR

08-SEP-2000; 2000US-0232080P.

PR PR

08-SEP-2000; 2000US-0232081P.

PR PR

12-SEP-2000; 2000US-0231968P.

PR PR

14-SEP-2000; 2000US-0232397P.

PR PR

14-SEP-2000; 2000US-0232398P.

PR PR

14-SEP-2000; 2000US-0232400P.

PR PR

14-SEP-2000; 2000US-0232401P.

PR PR

14-SEP-2000; 2000US-0233063P.

PR PR

14-SEP-2000; 2000US-0233064P.

PR PR

14-SEP-2000; 2000US-0233065P.

PR PR

21-SEP-2000; 2000US-0234223P.

PR PR

21-SEP-2000; 2000US-0234274P.

PR PR

25-SEP-2000; 2000US-0234997P.

PR PR

25-SEP-2000; 2000US-0234998P.

PR PR

26-SEP-2000; 2000US-0235484P.

PR PR

27-SEP-2000; 2000US-0235834P.

PR PR

27-SEP-2000; 2000US-0235836P.

PR PR

29-SEP-2000; 2000US-0236327P.

PR PR

29-SEP-2000; 2000US-0236367P.

PR PR

29-SEP-2000; 2000US-0236368P.

PR PR

29-SEP-2000; 2000US-0236369P.

PR PR

29-SEP-2000; 2000US-0236370P.

PR PR

02-OCT-2000; 2000US-0236802P.

PR PR

02-OCT-2000; 2000US-0237037P.

PR PR

02-OCT-2000; 2000US-0237038P.

PR PR

02-OCT-2000; 2000US-0237039P.

PR PR

02-OCT-2000; 2000US-0237040P.

PR PR

13-OCT-2000; 2000US-0239935P.

PR PR

13-OCT-2000; 2000US-0239937P.

PR PR

20-OCT-2000; 2000US-0240960P.

PR PR

20-OCT-2000; 2000US-0241221P.

PR PR

20-OCT-2000; 2000US-0241785P.

PR PR

20-OCT-2000; 2000US-0241786P.

PR PR

20-OCT-2000; 2000US-0241787P.

PR PR

20-OCT-2000; 2000US-0241808P.

PR PR

20-OCT-2000; 2000US-0241809P.

PR PR

20-OCT-2000; 2000US-0241826P.

PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 08-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Barash SC, Ruben SM;

DR WPI; 2001-457717/49.

XX Isolated pancreatic cancer polypeptide for treating, preventing and/or  
PT prognosing disorders related to the pancreas including pancreatic cancers  
PT and also for testing and detection e.g. diagnosis.

XX Disclosure; SEQ ID NO 589; 537pp; English.

XX The invention relates to an isolated polypeptide comprising an amino acid  
CC sequence at least 90% identical to 188 amino acid sequences fully defined  
CC in the specification and encoded by 188 cDNA clones fully defined in the  
CC specification. The invention also relates to a fragment having biological  
CC activity, a domain, an epitope, full length protein, variant, allelic  
CC variant or a species homologue of the fully defined sequence. The  
CC polynucleotide and polypeptide are useful for treating, preventing and/or  
CC prognosing disorders related to the pancreas including pancreatic cancer,  
CC pancreatitis, diabetes, endocrine disorders such as acromegaly or  
CC hyperthyroidism, and gastrointestinal disorders such as Crohn's disease  
CC and duodenal ulcers. The present sequence encodes a pancreatic cancer-

CC related polypeptide of the invention

XX SQ Sequence 14484 BP; 4086 A; 3663 C; 3382 G; 3353 T; 0 U; 0 Other;

Query Match 89.5%; Score 17; DB 4; Length 14484;

Best Local Similarity 100.0%; Pred. No. 2.3e-02;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATT 17

|||||

Db 142 CCCAGTGCATCCAAATT 158

RESULT 7

AAK89915

ID AAK89915 standard; DNA; 14484 BP.

XX AC AAK89915;

XX 05-NOV-2001 (first entry)

XX Human digestive system antigen genomic sequence SEQ ID NO: 3491.

XX Human; digestive system antigen; gene therapy; cancer; appendicitis;  
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;  
XX digestive system disorder; Meckel's diverticulum; ds.

XX Homo sapiens.

XX WO200155314-A2.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US001324.

XX 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205515P.

PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-0215135P.

PR 07-JUL-2000; 2000US-0216647P.

PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.

PR 11-JUL-2000; 2000US-0217496P.

PR 26-JUL-2000; 2000US-0220963P.

PR 26-JUL-2000; 2000US-0220964P.

PR 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0224519P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225214P.

PR 14-AUG-2000; 2000US-0225266P.

PR 14-AUG-2000; 2000US-0225267P.

PR 14-AUG-2000; 2000US-0225268P.

PR 14-AUG-2000; 2000US-0225270P.

PR 14-AUG-2000; 2000US-0225447P.

PR 14-AUG-2000; 2000US-0225477P.

PR 14-AUG-2000; 2000US-0225758P.

PR 14-AUG-2000; 2000US-0225759P.

PR 18-AUG-2000; 2000US-0226279P.

PR 22-AUG-2000; 2000US-0226681P.

PR 22-AUG-2000; 2000US-0226686P.

PR 22-AUG-2000; 2000US-0227182P.

PR 23-AUG-2000; 2000US-0227009P.

PR 30-AUG-2000; 2000US-0228924P.

PR 01-SEP-2000; 2000US-0229287P.

PR 01-SEP-2000; 2000US-0229343P.

```
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0232403P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-502630/55.
XX
XX Polynucleotides encoding digestive system antigens, useful for
PT diagnosing, treating, preventing and/or prognosing disorders of the
PT digestive system, particularly cancer and cancer metastases.
XX
XX Disclosure; SEQ ID NO 3491; 985pp; English.
XX
XX The present invention provides the protein and coding sequences of a
CC number of human digestive system antigens. These can be used in the
CC diagnosis, treatment and prevention of digestive system disorders,
CC including cancer, Meckel's diverticulum, bacterial or parasitic
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC ulcerative colitis. The present sequence is a genomic DNA fragment
CC encoding a digestive system antigen of the invention
XX
XX Sequence 14484 BP; 4086 A; 3663 C; 3382 G; 3353 T; 0 U; 0 Other;
SQ
Query Match 89.5%; Score 17; DB 4; Length 14484;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCCAGTGCATCCAAATT 17
Db 142 CCCAGTGCATCCAAATT 158
RESULT 8
ABA07272
ID ABA07272 standard; DNA; 14495 BP.
XX
XX ABA07272;
AC
XX
XX 14-JAN-2002 (first entry)
DT
XX
XX Human pancreatic cancer related genomic DNA, SEQ ID NO: 591.
XX
XX Human; cytostatic; antidiabetic; antiinflammatory; gastric; osteopathic;
XX antihormone; antiulcer; thyroid active; gene therapy; antisense therapy;
XX pancreatic cancer antigen inhibitor; pancreatic cancer; pancreatitis;
XX diabetes; endocrine disorder; acromegaly; hyperthyroidism;
XX gastrointestinal disorder; Crohn's disease; duodenal ulcer; ds.
XX
```

OS	Homo sapiens.		
XX	WO200155206-A1.	PR	27-SEP-2000; 2000US-0235834P.
XX		PR	27-SEP-2000; 2000US-0235836P.
XX		PR	29-SEP-2000; 2000US-0236327P.
XX		PR	29-SEP-2000; 2000US-0236367P.
PD		PR	29-SEP-2000; 2000US-0236368P.
XX	02-AUG-2001.	PR	29-SEP-2000; 2000US-0236369P.
XX		PR	29-SEP-2000; 2000US-0236370P.
PF	17-JAN-2001; 2001WO-US0001353.	PR	02-OCT-2000; 2000US-0236802P.
XX		PR	02-OCT-2000; 2000US-0237037P.
PR	31-JAN-2000; 2000US-0179065P.	PR	02-OCT-2000; 2000US-0237038P.
PR	04-FEB-2000; 2000US-0180628P.	PR	02-OCT-2000; 2000US-0237039P.
PR	24-FEB-2000; 2000US-0184664P.	PR	02-OCT-2000; 2000US-0237040P.
PR	02-MAR-2000; 2000US-0186350P.	PR	13-OCT-2000; 2000US-0239935P.
PR	16-MAR-2000; 2000US-0189874P.	PR	13-OCT-2000; 2000US-0239937P.
PR	17-MAR-2000; 2000US-0190076P.	PR	20-OCT-2000; 2000US-0240960P.
PR	18-APR-2000; 2000US-0198123P.	PR	20-OCT-2000; 2000US-0241221P.
PR	19-MAY-2000; 2000US-0205515P.	PR	20-OCT-2000; 2000US-0241785P.
PR	07-JUN-2000; 2000US-0209467P.	PR	20-OCT-2000; 2000US-0241786P.
PR	28-JUN-2000; 2000US-0214886P.	PR	20-OCT-2000; 2000US-0241787P.
PR	30-JUN-2000; 2000US-0215135P.	PR	20-OCT-2000; 2000US-0241808P.
PR	07-JUL-2000; 2000US-0216647P.	PR	20-OCT-2000; 2000US-0241809P.
PR	07-JUL-2000; 2000US-0216800P.	PR	20-OCT-2000; 2000US-0241826P.
PR	11-JUL-2000; 2000US-0217487P.	PR	01-NOV-2000; 2000US-0244617P.
PR	11-JUL-2000; 2000US-0217496P.	PR	08-NOV-2000; 2000US-0246474P.
PR	14-JUL-2000; 2000US-0218290P.	PR	08-NOV-2000; 2000US-0246475P.
PR	26-JUL-2000; 2000US-0220963P.	PR	08-NOV-2000; 2000US-0246476P.
PR	26-JUL-2000; 2000US-0220964P.	PR	08-NOV-2000; 2000US-0246477P.
PR	14-AUG-2000; 2000US-0224518P.	PR	08-NOV-2000; 2000US-0246478P.
PR	14-AUG-2000; 2000US-0224519P.	PR	08-NOV-2000; 2000US-0246523P.
PR	14-AUG-2000; 2000US-0225213P.	PR	08-NOV-2000; 2000US-0246524P.
PR	14-AUG-2000; 2000US-0225214P.	PR	08-NOV-2000; 2000US-0246525P.
PR	14-AUG-2000; 2000US-0225266P.	PR	08-NOV-2000; 2000US-0246526P.
PR	14-AUG-2000; 2000US-0225267P.	PR	08-NOV-2000; 2000US-0246527P.
PR	14-AUG-2000; 2000US-0225268P.	PR	08-NOV-2000; 2000US-0246528P.
PR	14-AUG-2000; 2000US-0225270P.	PR	08-NOV-2000; 2000US-0246532P.
PR	14-AUG-2000; 2000US-0225447P.	PR	08-NOV-2000; 2000US-0246609P.
PR	14-AUG-2000; 2000US-0225757P.	PR	08-NOV-2000; 2000US-0246610P.
PR	14-AUG-2000; 2000US-0225758P.	PR	08-NOV-2000; 2000US-0246611P.
PR	14-AUG-2000; 2000US-0225759P.	PR	08-NOV-2000; 2000US-0246613P.
PR	18-AUG-2000; 2000US-0226279P.	PR	17-NOV-2000; 2000US-0249207P.
PR	22-AUG-2000; 2000US-0226681P.	PR	17-NOV-2000; 2000US-0249208P.
PR	22-AUG-2000; 2000US-0226688P.	PR	17-NOV-2000; 2000US-0249209P.
PR	22-AUG-2000; 2000US-0227182P.	PR	17-NOV-2000; 2000US-0249210P.
PR	23-AUG-2000; 2000US-0227009P.	PR	17-NOV-2000; 2000US-0249211P.
PR	30-AUG-2000; 2000US-0228924P.	PR	17-NOV-2000; 2000US-0249212P.
PR	01-SEP-2000; 2000US-0229287P.	PR	17-NOV-2000; 2000US-0249213P.
PR	01-SEP-2000; 2000US-0229343P.	PR	17-NOV-2000; 2000US-0249214P.
PR	01-SEP-2000; 2000US-0229344P.	PR	17-NOV-2000; 2000US-0249215P.
PR	01-SEP-2000; 2000US-0229345P.	PR	17-NOV-2000; 2000US-0249216P.
PR	05-SEP-2000; 2000US-0229509P.	PR	17-NOV-2000; 2000US-0249217P.
PR	05-SEP-2000; 2000US-0229513P.	PR	17-NOV-2000; 2000US-0249218P.
PR	06-SEP-2000; 2000US-0230437P.	PR	17-NOV-2000; 2000US-0249219P.
PR	06-SEP-2000; 2000US-0230438P.	PR	17-NOV-2000; 2000US-0249244P.
PR	08-SEP-2000; 2000US-0231242P.	PR	17-NOV-2000; 2000US-0249245P.
PR	08-SEP-2000; 2000US-0231243P.	PR	17-NOV-2000; 2000US-0249264P.
PR	08-SEP-2000; 2000US-0231244P.	PR	17-NOV-2000; 2000US-0249265P.
PR	08-SEP-2000; 2000US-0231413P.	PR	17-NOV-2000; 2000US-0249297P.
PR	08-SEP-2000; 2000US-0231414P.	PR	17-NOV-2000; 2000US-0249299P.
PR	08-SEP-2000; 2000US-0232080P.	PR	01-DEC-2000; 2000US-0250160P.
PR	08-SEP-2000; 2000US-0232081P.	PR	01-DEC-2000; 2000US-0250391P.
PR	12-SEP-2000; 2000US-02331568P.	PR	05-DEC-2000; 2000US-0251030P.
PR	14-SEP-2000; 2000US-0232397P.	PR	05-DEC-2000; 2000US-0251988P.
PR	14-SEP-2000; 2000US-0232398P.	PR	05-DEC-2000; 2000US-0251989P.
PR	14-SEP-2000; 2000US-0232399P.	PR	06-DEC-2000; 2000US-0251479P.
PR	14-SEP-2000; 2000US-0232400P.	PR	08-DEC-2000; 2000US-0251856P.
PR	14-SEP-2000; 2000US-0232401P.	PR	08-DEC-2000; 2000US-0251868P.
PR	14-SEP-2000; 2000US-0233063P.	PR	08-DEC-2000; 2000US-0251869P.
PR	14-SEP-2000; 2000US-0233064P.	PR	08-DEC-2000; 2000US-0251989P.
PR	21-SEP-2000; 2000US-0234223P.	PR	08-DEC-2000; 2000US-0251990P.
PR	21-SEP-2000; 2000US-0234274P.	PR	11-DEC-2000; 2000US-0254097P.
PR	25-SEP-2000; 2000US-0234997P.	PR	05-JAN-2001; 2001US-0259678P.
PR	25-SEP-2000; 2000US-0234998P.	XX	
PR	26-SEP-2000; 2000US-0235484P.	XX	
		PA	(HUMA-) HUMAN GENOME SCI INC.

```
XX Rosen CA, Barash SC, Ruben SM;
PI WPI; 2001-45717/49.
XX
XX Isolated pancreatic cancer polypeptide for treating, preventing and/ or
PT prognosing disorders related to the pancreas including pancreatic cancers
PT and also for testing and detection e.g. diagnosis.
XX
XX Disclosure; SEQ ID NO 591; 537pp; English.
XX
XX The invention relates to an isolated polypeptide comprising an amino acid
CC sequence at least 90% identical to 188 amino acid sequences fully defined
CC in the specification and encoded by 188 cDNA clones fully defined in the
CC specification. The invention also relates to a fragment having biological
CC activity, a domain, an epitope, full length protein, variant, allelic
CC variant or a species homologue of the fully defined sequence. The
CC polynucleotide and polypeptide are useful for treating, preventing and/or
CC prognosing disorders related to the pancreas including pancreatic cancer,
CC pancreatitis, diabetes, endocrine disorders such as acromegaly or
CC hyperthyroidism, and gastrointestinal disorders such as Crohn's disease
CC and duodenal ulcers. The present sequence encodes a pancreatic cancer-
CC related polypeptide of the invention
XX
SQ Sequence 14495 BP; 4080 A; 3664 C; 3390 G; 3361 T; 0 U; 0 Other;
Query Match 89.5%; Score 17; DB 4; Length 14495;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCCAGTGCATCCAAATT 17
Db 142 CCCAGTGCATCCAAATT 158
RESULT 9
AAK89917
ID AAK89917 standard; DNA; 14495 BP.
XX
AC AAK89917;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human digestive system antigen genomic sequence SEQ ID NO: 3493.
XX
XX Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW digestive system disorder; Meckel's diverticulum; ds.
XX
XX Homo sapiens.
XX
XX WO200155314-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US0001324.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229511P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231411P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 12-SEP-2000; 2000US-0232081P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0244647P.
```

PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-502630/55.  
XX  
PT Polynucleotides encoding digestive system antigens, useful for  
PT diagnosing, treating, preventing and/or prognosing disorders of the  
PT digestive system, particularly cancer and cancer metastases.  
XX  
XX Disclosure; SEQ ID NO 3493; 986pp; English.  
XX  
CC The present invention provides the protein and coding sequences of a  
CC number of human digestive system antigens. These can be used in the  
CC diagnosis, treatment and prevention of digestive system disorders,  
CC including cancer, Meckel's diverticulum, bacterial or parasitic  
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or  
CC ulcerative colitis. The present sequence is a genomic DNA fragment  
CC encoding a digestive system antigen of the invention  
XX  
XX Sequence 14495 BP; 4080 A; 3664 C; 3390 G; 3361 T; 0 U; 0 Other;  
Query Match 89.5%; Score 17; DB 4; Length 14495;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATT 17  
Db 142 CCCAGTGCATCCAAATT 158  
RESULT 10  
AAH57251/c  
ID AAH57251 standard; cDNA; 359 BP.  
XX  
XX AAH57251;  
XX  
XX 10-SEP-2001 (first entry)  
XX Human liver specific cDNA sequence SEQ ID NO:91.  
XX Human; tissue specific; diagnosis; brain; heart; skeletal muscle; lung;  
XX liver; uterus; ovary; stomach; intestine; kidney; pancreas; es;  
XX metabolic disease; developmental disease; cytostatic; immunomodulatory;  
XX neuroprotective; gene therapy; cancer; immunopathology; neuropathology.  
XX Homo sapiens.  
XX  
XX WO200132927-A2.  
XX  
XX 10-MAY-2001.  
XX  
XX 02-NOV-2000; 2000WO-US030396.  
XX  
XX 04-NOV-1999; 99US-0163508P.  
XX (INCY-) INCYTE GENOMICS INC.  
XX  
XX Sornasse T, Seilhamer JJ, Watson GA;  
XX WPI; 2001-291057/30.  
XX  
XX New cell and tissue specific polynucleotides useful for diagnosis,  
XX prognosis or monitoring of treatments for disorders where the gene is  
XX associated with a cancer, immunopathology or neuropathology.  
XX  
XX Claim 1; Page 104-105; 327pp; English.  
XX  
XX AAH57161 to AAH57576 represent cell and tissue specific polynucleotide  
XX sequences (I). (I) can have cytostatic, immunomodulatory and  
XX neuroprotective activities, and can be used in gene therapy. (I) and  
XX proteins (II) encoded by them are used in high throughput screening  
XX assays to select DNA molecules, RNA molecules, peptide nucleic acids,  
XX mimetics, peptides, proteins, agonists, antagonists, antibodies or their  
XX fragments, immunoglobulins, inhibitors, drug compounds and pharmaceutical  
XX agents. Expression of (I) in a sample indicates the differentiation of  
XX embryonic stem cells into a tissue selected from brain, heart, kidney, used  
XX liver, lung, skeletal muscle or pancreatic tissues. (I) and (II) are used  
XX to produce an expression profile that defines a metabolic or  
XX developmental process, treatment, condition, disease or disorder. The  
XX gene profile can be used for diagnosis, prognosis or monitoring of  
XX treatments and for investigating a predisposition to a disorder where the  
XX gene is associated with a cancer, immunopathology or neuropathology  
XX  
XX Sequence 359 BP; 103 A; 68 C; 81 G; 93 T; 0 U; 14 Other;  
Query Match 86.3%; Score 16.4; DB 4; Length 359;  
Best Local Similarity 94.4%; Pred. No. 3.1e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 CCAGTGCATCCAAATTGA 19  
Db 220 CCAGTGCATCCAGATTGA 203  
RESULT 11  
ACL54336/c  
ID ACL54336 standard; cDNA; 430 BP.

```

XX AC ACL54336;
XX DT 24-MAR-2005 (first entry)
XX DE Human colon cancer differentially expressed polynucleotide, SEQ ID:471.
XX KW Differential expression; diagnosis; therapy; drug screening; cancer;
XX KW neoplasm; colon tumor; breast tumor; pancreas tumor; cytostatic; vaccine;
XX KW ss.
XX OS Homo sapiens.
XX PN WO200500087-A2.
XX PD 06-JAN-2005.
XX PF 13-MAY-2004; 2004WO-US015421.
XX PR 03-JUN-2003; 2003US-0475872P.
XX PA (CHIR ) CHIRON CORP.
XX PI Randazzo F, Moler E, Escobedo J, Garcia PD;
XX WPI; 2005-075421/08.
XX PT New isolated polynucleotides, which are differentially expressed in colon
XX PT cancer cell, useful for treating cancer, e.g. colon cancer, breast
XX PT cancer, or pancreatic cancer.
XX PS Claim 1; SEQ ID NO 471; 97pp; English.
XX CC The invention relates to 9672 polynucleotides (ACL53866-ACL63537) which
XX CC are differentially expressed in colon cancer cells. The invention also
XX CC relates to vectors and host cells comprising a differentially expressed
XX CC polynucleotide of the invention; a method for detecting a cancerous cell
XX CC by detection of a gene product of the polynucleotides; a method for
XX CC inhibiting a cancerous phenotype of a cell by inhibiting a gene product
XX CC of the polynucleotides; a method of treating an individual with cancer by
XX CC administration of a modulator of a gene product of the polynucleotides;
XX CC and an isolated antibody that specifically binds to a polypeptide encoded
XX CC by one of the 9672 polynucleotides. The polynucleotides, polypeptides,
XX CC antibodies, and methods are useful for the detection of cancerous cells;
XX CC for the diagnosis, prognosis and management of cancer; for the
XX CC identification of agents that modulate the phenotype of cancerous cells;
XX CC for the identification of therapeutic targets for cancer chemotherapy;
XX CC and for the treatment of cancer, especially colon cancer and metastasized
XX CC colon cancer, but also breast or pancreatic cancer. The polynucleotides
XX CC are also useful as a source of probes or primers for use in diagnostic
XX CC methods. The differentially expressed polynucleotides or their encoded
XX CC proteins can additionally be used as vaccines to modulate primary immune
XX CC responses for the prevention or treatment of cancer. The present sequence
XX CC represents a specifically claimed polynucleotide which is differentially
XX CC expressed in colon cancer. Note: The sequence data for this patent did
XX CC not form part of the printed specification, but was obtained in
XX CC electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 430 BP; 114 A; 83 C; 97 G; 136 T; 0 U; 0 Other;

Query Match 86.3%; Score 16.4; DB 14; Length 430;
Best Local Similarity 94.4%; Pred. No. 3.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATG 18
Db 214 CCCAGTGCATCCAAATG 197

RESULT 12
ABN93519
ID ABN93519 standard; DNA; 465 BP.

```

```

XX AC ABN93519;
XX DT 13-AUG-2002 (first entry)
XX DE Gene #17 used to diagnose liver cancer.
XX KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
XX KW metastatic liver tumor; cytostatic; expression profile; disease state;
XX KW disease progression; drug toxicity; drug efficacy; drug metabolism.
XX OS Homo sapiens.
XX PN WO200229103-A2.
XX PD 11-APR-2002.
XX PF 02-OCT-2001; 2001WO-US030589.
XX PR 02-OCT-2000; 2000US-0237054P.
XX PA (GENE-) GENE LOGIC INC.
XX PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX WPI; 2002-426119/45.
XX PT Diagnosing and detecting the progression of liver cancer, hepatocellular
XX PT carcinoma or metastatic liver tumor in a patient, involves detecting the
XX PT level of expression of two or more genes in a liver tissue sample.
XX PS Claim 1; SEQ ID NO 17; 298pp; English.
XX CC The invention relates to a novel method for diagnosing and detecting the
XX CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
XX CC tumor in a patient, and differentiating metastatic liver cancer from
XX CC hepatocellular carcinoma in a patient, involving detecting the level of
XX CC expression of two or more genes represented in ABN93503-ABN97455 in a
XX CC tissue sample. The method of the invention has hepatotropic, and
XX CC cytostatic activity. The method is useful for diagnosing and detecting
XX CC the progression of liver cancer, hepatocellular carcinoma and metastatic
XX CC liver carcinoma in a patient. The method is useful for identifying
XX CC expression profiles which serve as useful diagnostic markers as well as
XX CC markers that can be used to monitor disease states, disease progression,
XX CC drug toxicity, drug efficacy and drug metabolism. Note: The sequence data
XX CC for this patent did not form part of the printed specification, but was
XX CC obtained in electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 465 BP; 142 A; 83 C; 69 G; 171 T; 0 U; 0 Other;

Query Match 86.3%; Score 16.4; DB 6; Length 465;
Best Local Similarity 94.4%; Pred. No. 3.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCAGTGCATCCAAATG 19
Db 398 CCAGTGCATCCAGATG 415

RESULT 13
ABL28355/C
ID ABL28355 standard; DNA; 711 BP.
XX AC ABL28355;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 36538.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ds.

```



OS Drosophila melanogaster.  
 XX WO200171042-A2.  
 XX  
 XX  
 PD 27-SEP-2001.  
 XX  
 XX 23-MAR-2001; 2001WO-US009231.  
 XX  
 XX 23-MAR-2000; 2000US-0191637P.  
 PR 11-JUL-2000; 2000US-00614150.  
 XX  
 XX (PEKE ) PE CORP NY.  
 PA  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 PI WPI; 2001-656860/75.  
 XX  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signaling and cell-cell  
 PT interactions.  
 PT  
 XX Claim 1; SEQ ID NO 36538; 21pp + Sequence Listing; English.  
 PS  
 XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-  
 CC ABBS72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 711 BP; 186 A; 191 C; 192 G; 142 T; 0 U; 0 Other;  
 Query Match 86.3%; Score 16.4; DB 4; Length 711;  
 Best Local Similarity 94.4%; Pred. No. 3.3e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 CCAGTGCATCCAAATTGA 19  
 DB 334 CCAGTGCATCCAAATCGA 317  
 RESULT 14  
 ADF02970/C  
 ID ADF02970 standard; DNA; 1404 BP.  
 XX  
 XX ADF02970;  
 AC  
 XX 12-FEB-2004 (first entry)  
 DT  
 XX Bacterial polynucleotide #3255.  
 DE  
 XX Proteus mirabilis infection; bacterial infection; antibacterial;  
 KW immunostimulant; gene; ds.  
 KW  
 XX Proteus mirabilis.  
 OS  
 XX US6605709-B1.  
 PN  
 XX 12-AUG-2003.  
 PD  
 XX 05-APR-2000; 2000US-00543681.  
 PF  
 XX 09-APR-1999; 99US-0128706P.  
 PR  
 XX (GENO-) GENOME THERAPEUTICS CORP.  
 PA  
 XX Breton GL;  
 PI  
 XX WPI; 2003-895291/82.  
 DR

DR P-PSDB; ADF07142.  
 XX  
 XX New Proteus mirabilis polypeptides and polynucleotides, useful as  
 PT reagents for diagnosis of bacterial disease, as components of  
 PT antibacterial vaccines, as targets for antibacterial drugs, or as  
 PT biocontrol agents for plants.  
 XX  
 XX Disclosure; SEQ ID NO 3255; 870pp; English.  
 PS  
 XX The invention relates to new Proteus mirabilis polypeptides and  
 CC polynucleotides. The invention also relates to antibodies against the  
 CC polypeptides, methods for producing the polypeptides, a method of  
 CC generating vaccines for immunising an individual against P. mirabilis, a  
 CC method for evaluating a compound for the ability to bind a P. mirabilis  
 CC polypeptide and a method for screening test compounds for anti-bacterial  
 CC activity. The polypeptides and polynucleotides are useful as molecular  
 CC targets for diagnosing, preventing and treating pathological conditions  
 CC resulting from bacterial infection, as reagents for diagnosis of  
 CC bacterial diseases, as components of antibacterial vaccines, as targets  
 CC for antibacterial drugs or as bio-control agents for plants. This  
 CC sequence represents a Proteus mirabilis polynucleotide of the invention.  
 XX  
 SQ Sequence 1404 BP; 420 A; 299 C; 308 G; 377 T; 0 U; 0 Other;  
 Query Match 86.3%; Score 16.4; DB 10; Length 1404;  
 Best Local Similarity 94.4%; Pred. No. 3.5e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 CCCAGTGCATCCCAAAATTG 18  
 DB 824 CCCAGTGCAGCCCAAAATTG 807  
 RESULT 15  
 ADI22628/C  
 ID ADI22628 standard; DNA; 1721 BP.  
 XX  
 XX ADI22628;  
 AC  
 XX 22-APR-2004 (first entry)  
 DT  
 XX Human liver differentially expressed cDNA seq id 438.  
 DE  
 XX hepatotropic; antiinflammatory; virucide; cytostatic; gene therapy;  
 KW differentially gene expression; liver; toxin; liver disorder;  
 KW biliary cirrhosis; X-linked adrenoleukodystrophy; Zellweger syndrome;  
 KW hepatorenal syndrome; hepatitis; hepatocarcinoma; metabolic response;  
 KW toxicological response; ss; EST; expressed sequence tag; human.  
 XX  
 OS Homo sapiens.  
 OS  
 XX US2003165854-A1.  
 PN  
 XX 04-SEP-2003.  
 PD  
 XX 05-DEC-2001; 2001US-00006285.  
 PF  
 XX 05-DEC-2000; 2000US-0251986P.  
 PR  
 XX (CUNN/) CUNNINGHAM M J.  
 PA (KASE/) KASER M R.  
 XX  
 XX Cunningham MJ, Kaser MR;  
 PI  
 XX WPI; 2003-863697/80.  
 DR  
 XX New combination comprising a number of cDNAs that are differentially  
 PT expressed in a liver treated with a toxin, useful for diagnosing, staging  
 PT or treating liver disorders (e.g. biliary cirrhosis, hepatitis or  
 PT hepatocarcinoma).  
 XX  
 XX Claim 1; SEQ ID NO 438; 28pp; English.  
 PS  
 XX

CC The invention describes a combination comprising a number of cDNAs that  
CC are differentially expressed in a liver treated with a toxin and are  
CC selected from any of the 514 cDNAs listed in the specification, or their  
CC complements. The combination is useful in diagnosing, staging or treating  
CC a liver disorder (e.g. biliary cirrhosis, X-linked adrenoleukodystrophy,  
CC Zellweger syndrome, hepatocellular carcinoma, hepatitis or hepatocarcinoma),  
CC in monitoring diagnostic and therapeutic applications, in detecting  
CC metabolic and toxicological responses, and in elucidating drug mechanism  
CC of action. This sequence represents a cDNA differentially expressed in  
CC liver tissues in response to treatment with a toxin.  
XX  
SQ Sequence 1721 BP; 531 A; 314 C; 395 G; 481 T; 0 U; 0 Other;

Query Match 86.3%; Score 16.4; DB 10; Length 1721;  
Best Local Similarity 94.4%; Pred. No. 3.6e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCAGTGCATCCCAAAATTGA 19  
|||||  
DB 1112 CCAGTGCATCCCAAAATTGA 1095

RESULT 16  
AAH57459/c  
ID AAH57459 standard; cDNA; 1973 BP.

XX AC AAH57459;

XX DT 10-SEP-2001 (first entry)

XX DE Human liver cell specific cDNA sequence SEQ ID NO:299.

XX KW Human; tissue specific; diagnosis; brain; heart; skeletal muscle; lung;  
XX KW liver; uterus; ovary; stomach; intestine; kidney; pancreas; ss;  
XX KW metabolic disease; developmental disease; cytostatic; immunomodulatory;  
XX KW neuroprotective; gene therapy; cancer; immunopathology; neuropathology.  
XX OS Homo sapiens.

XX PN WO200132927-A2.

XX PD 10-MAY-2001.

XX PF 02-NOV-2000; 2000WO-US030396.

XX PR 04-NOV-1999; 99US-0163508P.

XX PA (INCY-) INCYTE GENOMICS INC.

XX PI Sornasse T, Seilhamer JU, Watson GA;

XX DR WPI; 2001-291057/30.

XX PT New cell and tissue specific polynucleotides useful for diagnosis,  
XX PT prognosis or monitoring of treatments for disorders where the gene is  
XX PT associated with a cancer, immunopathology or neuropathology.

XX PS Claim 1; Page 226; 327pp; English.

XX CC AAH57161 to AAH57576 represent cell and tissue specific polynucleotide  
XX CC sequences (I). (I) can have cytostatic, immunomodulatory and  
XX CC neuroprotective activities, and can be used in gene therapy. (I) and  
XX CC proteins (II) encoded by then are used in high throughput screening  
XX CC assays to select DNA molecules, RNA molecules, peptide nucleic acids,  
XX CC mimetics, peptides, proteins, agonists, antagonists, antibodies or their  
XX CC fragments, immunoglobulins, inhibitors, drug compounds and pharmaceutical  
XX CC agents. Expression of (I) in a sample indicates the differentiation of  
XX CC embryonic stem cells into a tissue selected from brain, heart, kidney,  
XX CC liver, lung, skeletal muscle or pancreatic tissues. (I) and (II) are used  
XX CC to produce an expression profile that defines a metabolic or  
XX CC developmental process, treatment, condition, disease or disorder. The  
XX CC gene profile can be used for diagnosis, prognosis or monitoring of  
XX CC treatments and for investigating a predisposition to a disorder where the

CC gene is associated with a cancer, immunopathology or neuropathology  
XX  
SQ Sequence 1973 BP; 623 A; 354 C; 440 G; 556 T; 0 U; 0 Other;  
Query Match 86.3%; Score 16.4; DB 4; Length 1973;  
Best Local Similarity 94.4%; Pred. No. 3.7e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCAGTGCATCCCAAAATTGA 19  
|||||

DB 1108 CCAGTGCATCCCAAAATTGA 1091

RESULT 17  
ABL28354  
ID ABL28354 standard; DNA; 2711 BP.

XX AC ABL28354;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 36535.

XX KW Drosophila; developmental biology; cell signalling; insecticide;

XX KW pharmaceutical; gene; ds.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US009231.

XX PR 23-MAR-2000; 2000US-0191637P.

XX PR 11-JUL-2000; 2000US-00614150.

XX PA (PEKE ) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more  
XX PT genes from Drosophila and for elucidating cell signaling and cell-cell  
XX PT interactions.

XX PS Claim 1; SEQ ID NO 36535; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent  
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is  
XX CC useful in developmental biology and in elucidating cell signalling and  
XX CC cell-cell interactions in higher eukaryotes for the development of  
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention  
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
XX CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB5773-  
XX CC ABB72072). The sequence data for this patent did not form part of the  
XX CC printed specification, but was obtained in electronic format directly  
XX CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 2711 BP; 678 A; 668 C; 684 G; 681 T; 0 U; 0 Other;

Query Match 86.3%; Score 16.4; DB 4; Length 2711;  
Best Local Similarity 94.4%; Pred. No. 3.8e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCAGTGCATCCCAAAATTGA 19  
|||||

DB 1378 CCAGTGCATCCCAAAATTGA 1395

RESULT 18  
ADC02494/c

```

ID ADC02494 standard; DNA; 2878 BP.
XX
AC ADC02494;
XX
DT 18-DEC-2003 (first entry)
XX
DE DNA encoding ubiquitin E2 #2.
XX
KW ubiquitin; Mdm2; p53; E2; ds; gene.
XX
OS Unidentified.
XX
FH Key Location/Qualifiers
FT CDS 45..542
FT /*tag= a
FT /product= "Ubiquitin E2"
XX
PN US2003104474-A1.
XX
PD 05-JUN-2003.
XX
PF 26-MAR-2002; 2002US-00108767.
XX
PR 03-APR-2000; 2000US-00542487.
PR 03-APR-2001; 2001US-00826312.
PR 04-MAR-2002; 2002US-00091139.
XX
XX (ISSA/) ISSAKANI S D.
PA (HUAN/) HUANG J.
PA (SHEU/) SHEUNG J.
PA (PRAY/) PRAY T R.
XX
PI Issakani SD, Huang J, Sheung J, Pray TR;
XX
XX WPI: 2003-787056/74.
XX P-PSDB; ADC02493.
XX
PT Assaying for candidate agent that modulates attachment of ubiquitin
PT moiety to Mdm2 protein comprises combining ubiquitin agent comprising
PT ubiquitin moiety, Mdm2 protein, and candidate agent.
XX
PS Disclosure; Fig 52B; 96pp; English.
XX
CC The present sequence represents a method of assaying for a candidate
CC agent that modulates the attachment of a ubiquitin moiety to an Mdm2
CC protein. The method is useful for assaying for a candidate agent that
CC modulates the attachment of a ubiquitin moiety to an Mdm2 or p53 protein.
CC The present sequence represents DNA encoding ubiquitin E2 #2.
XX
SQ Sequence 2878 BP; 744 A; 648 C; 755 G; 731 T; 0 U; 0 Other;
Query Match 86.3%; Score 16.4; DB 10; Length 2878;
Best Local Similarity 94.4%; Pred. No. 3.8e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CCCAGTGCATCCAAATTG 18
Db 1714 CCCAGTGCATCCAAATTG 1697
RESULT 19
ADF17879/c
ID ADF17879 standard; DNA; 2878 BP.
XX
AC ADF17879;
XX
DT 12-FEB-2004 (first entry)
XX
DE E2 ubiquitin conjugating agent UBC7 homologue DNA.
XX
KW UBC7; ds; gene; E2 ubiquitin conjugating agent;
KW ubiquitin-mediated proteolysis; high-throughput screening.
XX

```

---

```

OS Unidentified.
XX
XX Key Location/Qualifiers
FT CDS 45..542
FT /*tag= a
FT /product= "UBC7 homologue protein"
XX
PN US2003108947-A1.
XX
PD 12-JUN-2003.
XX
PF 20-MAY-2002; 2002US-00152156.
XX
PR 03-APR-2000; 2000US-00542497.
PR 03-APR-2001; 2001US-00826312.
PR 18-MAY-2001; 2001US-0291836P.
PR 04-MAR-2002; 2002US-00091139.
PR 04-MAR-2002; 2002US-00091139.
PR 26-MAR-2002; 2002US-00108767.
PR 26-MAR-2002; 2002US-00109460.
XX
XX (ISSA/) ISSAKANI S D.
PA (HUAN/) HUANG J.
PA (SHEU/) SHEUNG J.
PA (PRAY/) PRAY T R.
XX
PI Issakani SD, Huang J, Sheung J, Pray TR;
XX
XX WPI: 2003-810851/76.
XX P-PSDB; ADF17878.
XX
PT Assaying agent that modulates attachment of ubiquitin moiety to
PT ubiquitin agent, by combining first ubiquitin agent, candidate agent and
PT ubiquitin moiety, and assaying agent.
XX
PS Disclosure; Fig 52b; 95pp; English.
XX
CC This invention relates to a novel method for assaying ubiquitin agents
CC that are enzymatic components of ubiquitin-mediated proteolysis.
CC Specifically, it refers to a method for identifying ubiquitin-activating,
CC -conjugating and -ligating agents that work to modulate the attachment of
CC a ubiquitin moiety to at least one other ubiquitin agent. The method
CC comprises combining a first ubiquitin agent, candidate agent and
CC ubiquitin moiety, and subsequently assaying for the attachment of a
CC ubiquitin moiety to the first agent. The present invention does not
CC require a ubiquitin target protein and allows the analysis of many
CC different combinations of ubiquitin agents, without requiring the
CC identification of specific target proteins. Consequently, this method is
CC easy, applicable for high-throughput screening applications and so can
CC rapidly detect the activity of ubiquitin agents. This polynucleotide
CC sequence is the UBC7 homologue DNA, an E2 ubiquitin conjugating agent of
CC the invention.
XX
SQ Sequence 2878 BP; 744 A; 648 C; 755 G; 731 T; 0 U; 0 Other;
Query Match 86.3%; Score 16.4; DB 10; Length 2878;
Best Local Similarity 94.4%; Pred. No. 3.8e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CCCAGTGCATCCAAATTG 18
Db 1714 CCCAGTGCATCCAAATTG 1697
RESULT 20
ADQ85957/c
ID ADQ85957 standard; cDNA; 2878 BP.
XX
XX ADQ85957;
AC ADQ85957;
XX
DT 07-OCT-2004 (first entry)
XX
DE Human tumour-associated antigenic target (TAT) cDNA sequence #2829.

```

XX human; tumour-associated antigenic target; TAT; cytostatic; gene therapy;  
KW cancer; cell proliferative disorder; gene; ss.  
XX Homo sapiens.  
XX WO2004060270-A2.  
XX 22-JUL-2004.  
XX 15-OCT-2003; 2003WO-US029126.  
XX 18-OCT-2002; 2002US-0418988P.  
XX (GETH ) GENENTECH INC.  
PA (WUTD/) WU T D.  
PA (ZHOU/) ZHOU Y.  
XX Wu TD, Zhou Y;  
XX WPI; 2004-534300/51.  
XX New nucleic acid molecule and encoded polypeptide, for diagnosing,  
PT preventing or treating cell proliferative disorders such as cancer.  
XX Claim 1; SEQ ID NO 2829; 5504pp; English.  
XX The present invention describes an isolated tumour-associated antigenic  
CC target (TAT) nucleic acid comprising: (a) any of 4622 nucleotide  
CC sequences (see SEQ ID NO:1 to 4622); (b) the full-length coding region of  
CC (a); (c) the complement of (a) or (b); (d) a sequence that has 80%  
CC sequence identity to (a)-(c); or (e) a sequence that hybridises to (a)-  
CC (c). Also described: (1) an expression vector comprising the above  
CC nucleic acid; (2) a host cell comprising the above expression vector; (3)  
CC a process for producing a polypeptide; (4) an isolated polypeptide  
CC comprising: (a) an amino acid sequence encoded by any of the above  
CC nucleotide sequences; (b) an amino acid sequence encoded by the full-  
CC length coding region of the above nucleotide sequences; or (c) a sequence  
CC having at least 80% identical to (a) or (b); (5) a chimeric polypeptide  
CC comprising the above polypeptide fused to a heterologous polypeptide; (6)  
CC an isolated antibody that binds to the above polypeptide; (7) a process  
CC for producing the antibody; (8) an isolated oligopeptide that binds to  
CC the above polypeptide; (9) a tumour-associated antigenic target (TAT)  
CC binding organic molecule that binds to the above polypeptide; (10) a  
CC composition of matter comprising the above (chimeric) polypeptide,  
CC antibody, oligopeptide or TAT binding organic molecule, in combination  
CC with a carrier; (11) an article of manufacture comprising a container and  
CC the composition of matter contained within the container; (12) methods of  
CC inhibiting the growth of a cell that expresses the above protein, where  
CC the growth of the cell is at least in part dependent upon a growth  
CC potentiating effect of the above protein; (13) a method of  
CC therapeutically treating a mammal having a cancerous tumour comprising  
CC cells that express the above protein; (14) a method of determining the  
CC presence of a protein in a sample suspected of containing the protein  
CC described above; (15) methods of diagnosing the presence of a tumour in a  
CC mammal; (16) a method for treating or preventing a cell proliferative  
CC disorder associated with increased expression or activity of the above  
CC protein; and (17) a method of binding an antibody, oligopeptide or  
CC organic molecule to a cell that expresses the protein described above.  
CC The TAT sequences have cytostatic activities, and can be used in gene  
CC therapy. The composition and methods are useful for diagnosing,  
CC preventing or treating cancer. The composition is also used for preparing  
CC a medicament for the therapeutic treatment or diagnostic detection of a  
CC cell proliferative disorder or cancer. The present sequence represents a  
CC human TAT cDNA sequence from the present invention.  
XX SQ Sequence 2878 BP; 744 A; 648 C; 755 G; 731 T; 0 U; 0 Other;

Query Match 86.3%; Score 16.4; DB 12; Length 2878;  
Best Local Similarity 94.4%; Pred. No. 3.8e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CCCAGTGCATCCAAATG 18

Db 1714 CCCAGTGCATCCAAATG 1697  
RESULT 21  
ACN40857/c  
ID ACN40857 standard; cDNA; 2878 BP.  
XX AC ACN40857;  
XX 18-NOV-2004 (first entry)  
XX Tumour-associated antigenic target (TAT) cDNA DNA326904, SEQ ID NO:5917.  
XX Tumour-associated antigenic target; TAT; human; overexpression; cancer;  
KW tumour; diagnosis; cell proliferative disorder; breast cancer;  
KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;  
KW central nervous system cancer; bladder cancer; pancreatic cancer;  
KW cervical cancer; melanoma; leukaemia; hybridisation probe;  
KW chromosome identification; chromosome mapping; gene mapping;  
KW gene therapy; cytostatic; gene; ss.  
XX Homo sapiens.  
XX WO2004030615-A2.  
XX 15-APR-2004.  
XX 29-SEP-2003; 2003WO-US028547.  
XX 02-OCT-2002; 2002US-0414971P.  
XX (GETH ) GENENTECH INC.  
XX Wu TD, Zhang Z, Zhou Y;  
XX WPI; 2004-347921/32.  
XX P-PSDB; ABM82304.  
XX New tumour-associated antigenic target polypeptides and nucleic acids,  
PT useful in preparing a medicament for treating or detecting a  
PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or  
PT prostate cancer or tumor.  
XX Claim 1; SEQ ID NO 5917; 7273pp; English.  
XX The invention relates to human tumour-associated antigenic target (TAT)  
CC polypeptides, and their related nucleic acids. The TAT polypeptides are  
CC overexpressed in cancer tissues compared to normal tissues, and may thus  
CC serve as effective targets for the diagnosis and treatment of cancer in  
CC mammals. The invention also relates to nucleic acid and polypeptide  
CC sequences at least 80% identical to the TAT nucleic acids and  
CC polypeptides; expression vectors and host cells comprising a TAT nucleic  
CC acid; an antibody specific for a TAT polypeptide; a peptide or organic  
CC molecule which binds to a TAT polypeptide; fusion proteins comprising a  
CC TAT polypeptide; and methods and compositions for the treatment or  
CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,  
CC antibodies, antagonists, binding molecules and compositions are useful  
CC for diagnosing or treating a cell proliferative disorder associated with  
CC increased TAT expression, particularly cancers such as breast cancer,  
CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder  
CC cancer, pancreatic cancer, cervical cancer, cancers of the central  
CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be  
CC used as hybridisation probes, in chromosome and gene mapping, in  
CC chromosome identification and in gene therapy. The present sequence  
CC represents a TAT nucleic acid of the invention  
XX SQ Sequence 2878 BP; 744 A; 648 C; 755 G; 731 T; 0 U; 0 Other;

Query Match 86.3%; Score 16.4; DB 13; Length 2878;  
Best Local Similarity 94.4%; Pred. No. 3.8e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CCCAGTGCATCCAAATTG 18  
|||  
Db 1714 CCCAGTGCATCCAAATG 1697

RESULT 22  
ADB75604/c  
ID ADB75604 standard; cDNA: 2900 BP.

Query Match	86.3%	Score 16.4;	DB 10;	Length 2900;
Best Local Similarity	94.4%;	Pred. No. 3.8e+02;		
Matches 17; Conservative		0; Mismatches 1;	Indels 0;	Gaps 0;

RESULT 23  
ADZ49223/C  
ID ADZ49223 standard; DNA: 2919 BP.

XX		ADZ49223;	
XX	AC	XX	
XX	XX	XX	
DT	30-JUN-2005	(first entry)	
XX			
DE	Insulin signaling pathway related gene,	SEQ ID 552.	
XX			
KW	Antidiabetic; Gene Therapy; Non-insulin dependent diabetes;		
KW	insulin resistance; gene; ds.		
XX			
OS	Homo sapiens.		
XX			
PN	US2005085436-A1.		
XX			
PD	21-APR-2005.		
XX			
PB	08-JUL-2004; 2004US-00887553.		
XX			
PR	08-JUL-2003; 2003US-0485883P.		
XX			
PA	(LIHH/) LI H.		
PA	(MAJJ/) MA J.		
XX			
PI	Li H, Ma J;		
XX			
DR	WPI; 2005-305194/31.		
XX			
PT	Treating, preventing or ameliorating pathological conditions associated		
PT	with dysregulation of the insulin signaling pathway (ISP) comprises		
PT	administering to a subject an amount of a modulator of any of the		
PT	proteins regulated by ISP.		
XX			
PS	Disclosure; SEQ ID NO 552; 70pp; English.		
XX			
CC	The present invention relates to a method for treating, preventing or		
CC	ameliorating pathological conditions associated with dysregulation of the		
CC	insulin signaling pathway (ISP). The method comprises administering to a		
CC	subject a modulator for ISP-regulated proteins or a pharmaceutical		
CC	composition comprising the described modulator. The method is useful for		
CC	treating, preventing or ameliorating pathological conditions associated		
CC	with dysregulation of the ISP such as Type II diabetes or Type A syndrome		
CC	of insulin resistance. The present sequence is a human homolog of a		
CC	Drosophila gene regulated by ISP. Note: The sequence data for this patent		
CC	did not form part of the printed specification, but was obtained in		
CC	electronic format directly from USPTO at		
CC	seqdata.uspto.gov/sequence.html?DocID=20050085436.		
XX			
SQ	Sequence 2919 BP; 759 A; 657 C; 769 G; 734 T; 0 U; 0 Other;		
	Query Match	86.3%; Score 16.4; DB 14; Length 2919;	
	Best Local Similarity	94.4%; Pred. No. 3.8e+02;	
	Matches	17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
Oy	1	CCCAGTGCATCCAAATTG 18	
Db	1745	CCCAGTGCATCCAAATG 1728	
RESULT 24			
AAN91159/c			
ID	AAN91159 standard; DNA; 4740 BP.		
XX			
AC	AAN91159;		
XX			
DT	09-SEP-2004 (revised)		
DT	10-MAR-2003 (revised)		
DT	07-JUN-1990 (first entry)		
XX			
DE	pSK111 proteinase promoter fragment in S. cremoris SK211 related DNA..		
XX			
KW	Lactic acid bacteria; cheese; Streptococcus cremoris SK112; proteinase;		
XX	pSK111; ss.		
XX			



```

XX PF 30-JAN-2001; 2001WO-US0000669.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX XX WPI; 2001-483447/52.
XX DR Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human fetal liver.
XX PS Claim 1; SEQ ID NO 9371; 639pp + Sequence Listing; English.
XX CC The invention relates to a single exon nucleic acid probe for measuring
XX CC human gene expression in a sample derived from human foetal liver. The
XX CC single exon nucleic acid probes may be used for predicting, measuring and
XX CC displaying gene expression in samples derived from human fetal liver. The
XX CC present sequence is a single exon nucleic acid probe of the invention.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 512 BP; 124 A; 150 C; 123 G; 115 T; 0 U; 0 Other;

Query Match 83.2%; Score 15.8; DB 4; Length 512;
Best Local Similarity 89.5%; Pred. No. 6.4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATTGA 19
Db 318 CCCAGGGCTTCCAAATTGA 300

RESULT 27
AA140963/c
ID AA140963 standard; DNA; 512 BP.
XX AC AA140963;
XX DT 17-OCT-2001 (first entry)
XX DE Probe #9649 used to measure gene expression in human placenta sample.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder; ss.
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US0000663.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX XX WPI; 2001-483447/52.
XX DR Single exon nucleic acid probes for analyzing gene expression in human
XX PT hearts.
XX PS Claim 1; SEQ ID NO 9371; 639pp; English.
XX CC The present invention relates to single exon nucleic acid probes for
XX CC measuring human gene expression in a sample derived from human heart. The
XX CC present sequence is one such probe. The probes may be used for
XX CC predicting, measuring and displaying gene expression in samples derived
XX CC from the human heart via microarrays. By measuring gene expression, the
XX CC probes are useful for predicting, diagnosing, grading, staging,

```

```

PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX XX WPI; 2001-488897/53.
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human placenta.
XX PS Claim 25; SEQ ID NO 9649; 654pp; English.
XX CC The present invention relates to single exon nucleic acid probes (SENP).
XX CC The present sequence is one such probe. The probes are useful for
XX CC producing a microarray for predicting, measuring and displaying gene
XX CC expression in samples derived from human placenta. The probes are useful
XX CC for antenatal diagnosis of human genetic disorders
XX SQ Sequence 512 BP; 124 A; 150 C; 123 G; 115 T; 0 U; 0 Other;

Query Match 83.2%; Score 15.8; DB 4; Length 512;
Best Local Similarity 89.5%; Pred. No. 6.4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATTGA 19
Db 318 CCCAGGGCTTCCAAATTGA 300

RESULT 28
ABA28970/c
ID ABA28970 standard; DNA; 512 BP.
XX AC ABA28970;
XX DT 23-JAN-2002 (first entry)
XX DE Probe #7436 for gene expression analysis in human heart cell sample.
XX KW Human; gene expression; heart; microarray; vascular system; probe;
XX KW cardiovascular disease; hypertension; cardiac arrhythmia;
XX KW congenital heart disease; ss.
XX OS Homo sapiens.
XX PN WO200157274-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US0000666.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX XX WPI; 2001-488899/53.
XX DR Single exon nucleic acid probes for analyzing gene expression in human
XX PT hearts.
XX PS Claim 1; SEQ ID NO 7436; 530pp; English.
XX CC The present invention relates to single exon nucleic acid probes for
XX CC measuring human gene expression in a sample derived from human heart. The
XX CC present sequence is one such probe. The probes may be used for
XX CC predicting, measuring and displaying gene expression in samples derived
XX CC from the human heart via microarrays. By measuring gene expression, the
XX CC probes are useful for predicting, diagnosing, grading, staging,

```

CC monitoring and prognosing diseases of the human heart and vascular system  
 CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and  
 CC congenital heart disease. Note: The sequence data for this patent did not  
 CC form part of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 512 BP; 124 A; 150 C; 123 G; 115 T; 0 U; 0 Other;

Query Match 83.2%; Score 15.8; DB 4; Length 512;  
 Best Local Similarity 89.5%; Pred. No. 6.4e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CCCAGTGCATCCAAATTGA 19  
 ||||| ||||| ||||| |||||  
 Db 318 CCCAGGGCTTCCAAATTGA 300

## RESULT 29

AAK35248/c  
 ID AAK35248 standard; DNA; 512 BP.

XX  
 AC AAK35248;

DT 06-NOV-2001 (first entry)

DE Human bone marrow expressed single exon probe SEQ ID NO: 9805.

XX Human; bone marrow expressed exon; gene expression analysis; probe;  
 KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.

XX Homo sapiens.

XX WO200157276-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000668.

XX 04-FEB-2000; 2000US-0180312P.

PR 26-MAY-2000; 2000US-0207456P.

PR 30-JUN-2000; 2000US-00608408.

PR 03-AUG-2000; 2000US-00632366.

PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.

PR 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488900/53.

XX Human genome-derived single exon nucleic acid probes useful for analyzing  
 PT gene expression in human bone marrow.

XX Example 4; SEQ ID NO 9805; 658pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC bone marrow. They can be used to measure gene expression in bone marrow  
 CC samples, which may enable the improved diagnosis and treatment of cancers  
 CC such as lymphoma, leukaemia and myeloma. The present sequence is one of  
 CC the probes of the invention

XX Sequence 512 BP; 124 A; 150 C; 123 G; 115 T; 0 U; 0 Other;

Query Match 83.2%; Score 15.8; DB 4; Length 512;  
 Best Local Similarity 89.5%; Pred. No. 6.4e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CCCAGTGCATCCAAATTGA 19  
 ||||| ||||| ||||| |||||  
 Db 318 CCCAGGGCTTCCAAATTGA 300

## RESULT 30

AAK09359/c  
 ID AAK09359 standard; DNA; 512 BP.

XX  
 AC AAK09359;

DT 05-NOV-2001 (first entry)

XX Human brain expressed single exon probe SEQ ID NO: 9350.

XX Human; brain expressed exon; gene expression analysis; probe; microarray;  
 KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer;  
 KW ss.

XX Homo sapiens.

XX WO200157275-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000667.

XX 04-FEB-2000; 2000US-0180312P.

PR 26-MAY-2000; 2000US-0207456P.

PR 30-JUN-2000; 2000US-00608408.

PR 03-AUG-2000; 2000US-00632366.

PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.

PR 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483446/52.

XX Single exon nucleic acid probes for analyzing gene expression in human  
 PT brains.

XX Example 4; SEQ ID NO 9350; 650pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC brain. They can be used to measure gene expression in brain cell samples,  
 CC which may enable the diagnosis and improved treatment of nervous system  
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
 CC epilepsy and cancers. The present sequence is one of the probes of the  
 CC invention

XX Sequence 512 BP; 124 A; 150 C; 123 G; 115 T; 0 U; 0 Other;

Query Match 83.2%; Score 15.8; DB 4; Length 512;  
 Best Local Similarity 89.5%; Pred. No. 6.4e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CCCAGTGCATCCAAATTGA 19  
 ||||| ||||| ||||| |||||  
 Db 318 CCCAGGGCTTCCAAATTGA 300

## RESULT 31

ABS34992/c  
 ID ABS34992 standard; DNA; 512 BP.

XX  
 AC ABS34992;

DT 25-FEB-2003 (first entry)

XX Human liver single exon probe, SEQ ID No 9982.

XX Human; single exon nucleic acid probe; liver; cirrhosis;



KW hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;  
 KW coronary heart disease; ss.

XX Homo sapiens.

XX WO200157273-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000564.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234687P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488899/53.

XX Human genome-derived single exon nucleic acid probes useful for analyzing  
 PT gene expression in human adult liver.

XX Claim 1; SEQ ID NO 9982; 658pp; English.

XX The invention relates to a single exon nucleic acid probe (SENP) (I) for  
 CC measuring human gene expression in a sample derived from human adult  
 CC liver, comprising one of 13109 defined nucleotide sequences given in the  
 CC specification (or complements/ fragments). The probe hybridises at high  
 CC stringency to a nucleic acid molecule expressed in the human adult liver.  
 CC (I) may be used for predicting, measuring and displaying gene expression  
 CC in samples derived from human adult liver. The genes identified may be  
 CC involved in genetic liver diseases such as cirrhosis,  
 CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is  
 CC associated with coronary heart disease. ABS25011-AB551005 represent human  
 CC liver single exon nucleic acid probes of the invention. Note: The  
 CC sequence information for this patent does not appear in the printed  
 CC specification but was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 512 BP; 124 A; 150 C; 123 G; 115 T; 0 U; 0 Other;

Query Match 83.2%; Score 15.8; DB 4; Length 512;

Best Local Similarity 89.5%; Pred. No. 6.4e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATTGA 19

Db 318 CCCAGGCTTCCAAATTGA 300

RESULT 32

ABS09663/c

ID ABS09663 standard; DNA; 512 BP.

XX ABS09663;

XX 19-AUG-2002 (first entry)

XX Human genome-derived single exon probe from lung SEQ ID No 9654.

XX Human; db; single exon probe; asthma; lung cancer; COPD; ILD;  
 KW chronic obstructive pulmonary disease; interstitial lung disease;  
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;  
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;

KW primary ciliary dyskinesia; pulmonary hypertension;  
 KW hyaline membrane disease.

XX Homo sapiens.

XX WO200186003-A2.

XX 15-NOV-2001.

XX 30-JAN-2001; 2001WO-US000665.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234687P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2002-114183/15.

XX Spatially-addressable set of single exon nucleic acid probes, used to  
 PT measure gene expression in human lung samples.

XX Claim 1; SEQ ID NO 9654; 634pp; English.

XX The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human lung comprising single exon nucleic acid probes having one of  
 CC 12614 nucleic acid sequences mentioned in the specification, or their  
 CC complements or the 12387 open reading frames derived from the 12614  
 CC probes. Also included are a microarray comprising the novel set of probes  
 CC; the novel set of probes which hybridise at high stringency to a nucleic  
 CC acid expressed in the human lung; measuring gene expression in a sample  
 CC derived from human lung, comprising (a) contacting the array with a  
 CC collection of detectably labeled nucleic acids derived from human lung  
 CC mRNA, and (b) measuring the label detectably bound to each probe of the  
 CC array; identifying exons in a eukaryotic genome, comprising (a)  
 CC algorithmically predicting at least one exon from genomic sequences of  
 CC the eukaryote; and (b) detecting specific hybridisation of detectably  
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
 CC having a fragment identical to the predicted exon, the probe is included  
 CC in the above mentioned microarray; assigning exons to a single gene,  
 CC comprising (a) identifying exons from genomic sequence by the method  
 CC above and (b) measuring the expression of each of the exons in several  
 CC tissues and/or cell types using hybridisation to a single exon  
 CC microarrays having a probe with the exon, where a common pattern of  
 CC expression of the exons in the tissues and/or cell types indicates that  
 CC the exons should be assigned to a single gene; a peptide comprising one  
 CC of 12011 sequences, mentioned in the specification, or encoded by the  
 CC probes/open reading frames (ORF). The probes are used for gene expression  
 CC analysis, and for identifying exons in a gene, particularly using human  
 CC lung derived mRNA and for the study of lung diseases such as asthma, lung  
 CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung  
 CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,  
 CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-  
 CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary  
 CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,  
 CC Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary  
 CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The  
 CC present sequence is a single exon probe of the invention. Note: The  
 CC sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 512 BP; 124 A; 150 C; 123 G; 115 T; 0 U; 0 Other;

Query Match 83.2%; Score 15.8; DB 6; Length 512;

Best Local Similarity 89.5%; Pred. No. 6.4e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATTGA 19  
||||| ||| ||||| |||||  
Db 318 CCCAGGCTTCCAAATTGA 300

RESULT 33  
AAH15533/C  
ID AAH15533 standard; cDNA; 1998 BP.  
AC AAH15533;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human cDNA sequence SEQ ID NO:13804.  
XX  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
FN EF1074617-A2.  
XX  
PD 07-FEB-2001.  
XX  
PF 28-JUL-2000; 2000EP-00116126.  
XX  
PR 29-JUL-1999; 99JP-00248036.  
PR 27-AUG-1999; 99JP-00300253.  
PR 11-JAN-2000; 2000JP-00118776.  
PR 02-MAY-2000; 2000JP-00183767.  
PR 09-JUN-2000; 2000JP-00241899.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Ota T, Isogai T, Nishikawa T, Hayaashi K, Saito K, Yamamoto J;  
PI Iehii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
DR WPI; 2001-318749/34.  
XX  
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.  
XX  
PS Claim 8; SEQ ID NO 13804; 2537pp + Sequence Listing; English.  
XX  
CC The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly, full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialized methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention  
XX  
SQ Sequence 1998 BP; 521 A; 353 C; 474 G; 650 T; 0 U; 0 Other;

Query Match 83.2%; Score 15.8; DB 4; Length 1998;

Best Local Similarity 89.5%; Pred. No. 7.4e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATTGA 19  
||||| ||| ||||| |||||  
Db 777 CCCATGCATCTAAATTGA 759

RESULT 34  
ADZ36079  
ID ADZ36079 standard; cDNA; 2538 BP.  
XX  
AC ADZ36079;  
XX  
DT 14-JUL-2005 (first entry)  
XX  
DE Human NFKB1sv2 encoding cDNA SEQ ID NO:5.  
XX  
KW nuclear factor kappa B; chromosome-4; gene; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..2538  
FT /\*tag= a  
FT /product= "NFKB1sv2"  
XX  
PN US2005089912-A1.  
XX  
PD 28-APR-2005.  
XX  
PF 06-OCT-2004; 2004US-00960409.  
XX  
PR 07-OCT-2003; 2003US-0509361P.  
XX  
PA (ARMO/) ARMOUR C D.  
PA (CAST/) CASTLE J C.  
PA (GARR/) GARRETT-ENGELE P W.  
PA (KANZ/) KAN Z.  
PA (RAYM/) RAYMOND C K.  
PA (TSIN/) TSINOREMAS N F.  
XX  
PI Armour CD, Castle JC, Garrett-Engеле PW, Kan Z, Raymond CK;  
PI Tsinoremas NF;  
XX  
DR WPI; 2005-344276/35.  
DR P-PSDB; ADZ36080.  
XX  
PT New purified nuclear factor kappa-B sub unit 1 polypeptide, useful for screening for a compound capable of binding to or interacting with NFKB1sv2 protein or its fragment.  
XX  
PS Claim 1; SEQ ID NO 5; 27pp; English.  
XX  
CC The invention relates to a purified nuclear factor kappa B subunit 1 (NFKB1) polypeptide (I) comprising the 845 amino acid sequence of ADZ36080. Also described: (1) a purified human NFKB1 nucleic acid comprising the 2538 nucleotide sequence of ADZ36079, or its complement; (2) an expression vector comprising a nucleotide sequence encoding ADZ36080, where the nucleotide sequence is transcriptionally coupled to an exogenous promoter; (3) screening for a compound capable of binding to or NFKB1sv2; and (4) screening for a compound capable of binding to or interacting with a NFKB1sv2 protein or its fragment. (I) is useful for screening for a compound capable of binding to or interacting with NFKB1sv2 protein or its fragment. The present sequence encodes the human NFKB1 isoform splice variant 2 (NFKB1sv2), which is used in an example from the present invention. The human NFKB1 gene is located on chromosome 4, more specifically to region 4q24.  
XX  
SQ Sequence 2538 BP; 737 A; 564 C; 653 G; 584 T; 0 U; 0 Other;

Query Match 83.2%; Score 15.8; DB 14; Length 2538;

Best Local Similarity 89.5%; Pred. No. 7.6e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATTGA 19  
|||||  
Db 736 CCCAATGCATCCAACTTGA 754

## RESULT 35

ADZ36077  
ID ADZ36077 standard; cDNA; 2811 BP.

XX AC ADZ36077;

XX DT 14-JUL-2005 (first entry)

XX DE Human NFKB1sv1 encoding cDNA SEQ ID NO:3.

XX KW nuclear factor kappa B; chromosome-4; gene; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT CDS 1..2811

FT /\*tag= a

FT /product= "NFKB1sv1"

XX US2005089912-A1.

XX PD 28-APR-2005.

XX PF 06-OCT-2004; 2004US-00960409.

XX PR 07-OCT-2003; 2003US-0509361P.

XX PA (ARMO/) ARMOUR C D.

XX PA (CAST/) CASTLE J C.

XX PA (GARR/) GARRETT-ENGLE P W.

XX PA (KANZ/) KAN Z.

XX PA (RAYM/) RAYMOND C K.

XX PA (TSIN/) TSINOREMAS N F.

XX PI Armour CD, Castle JC, Garrett-Engle PW, Kan Z, Raymond CK;

XX PI Tsinoremas NF;

XX DR WPI; 2005-344276/35.

XX DR P-PSDB; ADZ36078.

XX PT New purified nuclear factor kappa-B sub unit 1 polypeptide, useful for

PT screening for a compound capable of binding to or interacting with

PT NFKB1sv2 protein or its fragment.

XX PS Example 3; SEQ ID NO 3; 27pp; English.

XX CC The invention relates to a purified nuclear factor kappa B subunit 1

CC (NFKB1) polypeptide (1) comprising the 845 amino acid sequence of

CC ADZ36080. Also described: (1) a purified human NFKB1 nucleic acid

CC comprising the 2538 nucleotide sequence of ADZ36079, or its complement;

CC (2) an expression vector comprising a nucleotide sequence encoding

CC ADZ36080, where the nucleotide sequence is transcriptionally coupled to

CC an exogenous promoter; (3) screening for a compound capable of binding to

CC NFKB1sv2; and (4) screening for a compound capable of binding to or

CC interacting with a NFKB1sv2 protein or its fragment. (1) is useful for

CC screening for a compound capable of binding to or interacting with

CC NFKB1sv2 protein or its fragment. The present sequence encodes the human

CC NFKB1 isoform splice variant 1 (NFKB1sv1), which is used in an example

CC from the present invention. The human NFKB1 gene is located on chromosome

CC 4, more specifically to region 4q24.

XX SQ Sequence 2811 BP; 794 A; 641 C; 736 G; 640 T; 0 U; 0 Other;

## Query Match

Best Local Similarity 83.2%; Score 15.8; DB 14; Length 2811;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATTGA 19  
|||||  
Db 637 CCCAATGCATCCAACTTGA 655

## RESULT 36

ADA02614

ID ADA02614 standard; cDNA; 2907 BP.

XX AC ADA02614;

XX DT 06-NOV-2003 (first entry)

XX DE Human NFKB1 carcinoma associated coding sequence, SEQ ID NO:1132.

XX KW Human; carcinoma associated; oncogene; carcinoma; cancer; breast;

XX KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;

XX KW gene; ss.

XX OS Homo sapiens.

XX PN WO2003057146-A2.

XX PD 17-JUL-2003.

XX PF 26-DEC-2002; 2002WO-US041414.

XX PR 26-DEC-2001; 2001US-00035832.

XX PA (SAGR-) SAGRES DISCOVERY.

XX PI Morris DW;

XX WPI; 2003-587068/55.

XX PT New recombinant nucleic acid encoding carcinoma associated protein,

XX PT useful for preparing compositions for treating carcinomas.

XX PS Claim 1; SEQ ID NO 1132; 245pp; English.

XX CC The invention relates to recombinant carcinoma associated (CA) nucleic

XX CC acid sequences from mouse and human (ADA01482-ADA03094), and to

XX CC recombinant carcinoma associated proteins (CAP) encoded by them. The

XX CC invention also encompasses expression vectors and host cells comprising a

XX CC CA nucleic acid, a polypeptide (especially an antibody) that specifically

XX CC binds to the protein, and a biochip comprising CA nucleic acid or

XX CC fragments thereof. The sequences of the invention were identified using

XX CC oncogenic retroviruses, which insert into the genome of the host organism

XX CC at random. Many of these do not carry transduced host oncogenes or

XX CC pathogenic trans-acting viral genes, meaning that cancer incidence is a

XX CC direct consequence of the effects of proviral integration into host

XX CC protooncogenes. The CA nucleic acid sequences can be used to diagnose

XX CC carcinoma (especially breast cancer, prostate cancer, lymphoma or

XX CC leukaemia) or a propensity to carcinoma by determination of the sequence

XX CC of a CA gene, or by determination of CA gene expression in particular

XX CC tissues. CA nucleic acids, proteins and antibodies are also useful as

XX CC therapeutic agents and in screening and evaluating drug candidates. The

XX CC present sequence represents a specifically claimed human CA nucleic acid

XX CC sequence of the invention. Note: The complete sequence data for this

XX CC patent did not form part of the printed specification, but was obtained

XX CC in electronic format directly from WIPO at

XX CC ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 2907 BP; 820 A; 662 C; 759 G; 666 T; 0 U; 0 Other;

XX PT Query Match 83.2%; Score 15.8; DB 9; Length 2907;

XX PT Best Local Similarity 89.5%; Pred. No. 7.7e+02;

XX PT Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATTGA 19

|||||

Db 733 CCCAATGCATCCAACTTGA 751

```

RESULT 37
ADB72352
ID ADB72352 standard; cDNA; 2907 BP.
XX
XX AC ADB72352;
XX
XX DT 04-DEC-2003 (first entry)
XX
XX DE Human NFKB1 cDNA.
XX
XX KW human; ss; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
XX KW cancer; neoplasm; adenocarcinoma; sarcoma; gene.
XX
XX OS Homo sapiens.
XX
XX PN WO2003008583-A2.
XX
XX PD 30-JAN-2003.
XX
XX PF 26-DEC-2001; 2001WO-US051291.
XX
XX PR 02-MAR-2001; 2001US-00798586.
XX PR 23-OCT-2001; 2001US-00004113.
XX PR 08-NOV-2001; 2001US-00052482.
XX PR 30-NOV-2001; 2001US-00997722.
XX PR 20-DEC-2001; 2001US-00034650.
XX
XX PA (SAGR-) SAGRES DISCOVERY.
XX
XX PI Morris DW, Engelhard EK;
XX
XX WI WI; 2003-239337/23.
XX
XX PT New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
XX PT cancers, neoplasm, adenocarcinoma, or sarcomas.
XX
XX PS Claim 1; SEQ ID NO 180; 2304pp; English.
XX
XX CC The invention relates to a novel recombinant nucleic acid comprising a
XX CC nucleotide sequence selected from any of the 660 sequences fully defined
XX CC in the specification. A polynucleotide of the invention has cytostatic
XX CC activity, and may have a use in gene therapy, or in a vaccine. The
XX CC recombinant nucleic acids and polypeptides are useful for treating
XX CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
XX CC sarcomas. The present sequence represents a human cDNA of the invention.
XX
XX SQ Sequence 2907 BP; 820 A; 662 C; 759 G; 666 T; 0 U; 0 Other;

Query Match 83.2%; Score 15.8; DB 10; Length 2907;
Best Local Similarity 89.5%; Pred. No. 7.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATTGA 19
Db 733 CCCAATGCATCCAACTTGA 751

RESULT 38
ADE82946
ID ADE82946 standard; DNA; 2907 BP.
XX
XX AC ADE82946;
XX
XX DT 29-JAN-2004 (first entry)
XX
XX DE Human NFKB1 coding sequence.
XX
XX KW human; cancer-associated nucleic acid; screening; cancer; lymphoma;
XX KW leukaemia; breast cancer; gene therapy; vaccine; ds.
XX
XX OS Homo sapiens.

```

```

XX PN WO2003080808-A2.
XX
XX PD 02-OCT-2003.
XX
XX PF 21-MAR-2003; 2003WO-US008919.
XX
XX PR 21-MAR-2002; 2002US-0367025P.
XX
XX PA (SAGR-) SAGRES DISCOVERY.
XX
XX PI Morris DW;
XX
XX WI WI; 2003-865119/80.
XX
XX PT New cancer-associated proteins and nucleic acids, useful for screening
XX PT for anticancer activity in a potential drug, or for detecting,
XX PT diagnosing, preventing and treating cancers, e.g. lymphoma, leukemia or
XX PT breast cancer.
XX
XX PS Claim 18; SEQ ID NO 30; 248pp; English.
XX
XX CC The invention comprises human and mouse cancer-associated nucleic acid
XX CC sequences. The cancer associated nucleic acids of the invention are
XX CC useful for screening for anticancer activity in a potential drug, as well
XX CC as detecting, diagnosing, preventing and treating cancers (e.g. lymphoma,
XX CC leukaemia, or breast cancer). The present sequence represents a cancer-
XX CC associated nucleic acid of the invention.
XX
XX SQ Sequence 2907 BP; 820 A; 662 C; 759 G; 666 T; 0 U; 0 Other;

Query Match 83.2%; Score 15.8; DB 10; Length 2907;
Best Local Similarity 89.5%; Pred. No. 7.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATTGA 19
Db 733 CCCAATGCATCCAACTTGA 751

RESULT 39
ADE95862
ID ADE95862 standard; DNA; 2907 BP.
XX
XX AC ADE95862;
XX
XX DT 12-FEB-2004 (first entry)
XX
XX DE Human NFKB1 gene coding sequence.
XX
XX KW cancer diagnosis; cancer treatment; carcinoma; cytostatic; gene therapy;
XX KW lymphoma; breast cancer; prostate cancer; leukaemia; ds; human; NFKB1.
XX
XX OS Homo sapiens.
XX
XX PN WO2003039484-A2.
XX
XX PD 15-MAY-2003.
XX
XX PF 08-NOV-2002; 2002WO-US036071.
XX
XX PR 08-NOV-2001; 2001US-00052482.
XX
XX PA (SAGR-) SAGRES DISCOVERY.
XX
XX PI Morris DW, Engelhard EK;
XX
XX WI WI; 2003-441462/41.
XX
XX PT New carcinoma associated nucleic acids and proteins, useful for screening
XX PT drug candidates, or for diagnosing and treating carcinomas, e.g.
XX PT lymphoma, breast cancer, prostate cancer or leukemia.
XX
XX

```

PS Claim 1; SEQ ID NO 120; 793pp; English.

CC This invention relates to novel recombinant nucleic acids for use in  
CC diagnosis and treatment of cancer, especially carcinomas, as well as the  
CC use of compositions in screening methods. The compositions of the  
CC invention may have cytostatic activity whilst the disclosed sequences may  
CC be useful for gene therapy. The carcinoma associated nucleic acids and  
CC proteins are useful for diagnosing and treating carcinomas, for example  
CC lymphoma, breast cancer, prostate cancer or leukaemia, or for screening  
CC drug candidates or bioactive agents capable of binding to, or modulating  
CC the activity of, a carcinoma associated protein. The present sequence is  
CC the coding DNA sequence of the human NFKB1 gene which is a carcinoma  
CC associated gene of the invention.

XX  
SQ Sequence 2907 BP; 820 A; 662 C; 759 G; 666 T; 0 U; 0 Other;

Query Match 83.2%; Score 15.8; DB 10; Length 2907;

Best Local Similarity 89.5%; Pred. No. 7.7e+02; Mismatches 2; Indels 0; Gaps 0;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCCAATTGA 19

|||||

Db 733 CCCAATGCATCCCAACTTGA 751

RESULT 40

ADV42551

ID ADV42551 standard; cDNA; 2907 BP.

XX

AC ADV42551;

XX

DT 10-MAR-2005 (first entry)

XX

DE Human psychoneuroendocrinimmune expressed sequence tag SEQ ID NO 179.

XX

KW microarray; psychoneuroendocrinimmune; chronic fatigue;  
KW non-insulin dependent diabetes; allergy; immune disorder; inflammation;  
KW cancer; neoplasm; infection; expressed sequence tag; ss.

OS Homo sapiens.

XX

PN WO2004108899-A2.

XX

PD 16-DEC-2004.

XX

PF 04-JUN-2004; 2004WO-US017686.

XX

PR 04-JUN-2003; 2003US-0475915P.

XX

PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.

XX

PI Nicholson A, Vernon SD;

XX

DR WPI; 2005-031682/03.

XX

PT New microarray comprising probes for genes involved in  
PT psychoneuroendocrinimmune (PNI) activity, useful in diagnosing a  
PT condition associated with PNI activity, e.g., inflammatory or infectious  
PT diseases.

XX

PS Claim 1; SEQ ID NO 179; 254pp; English.

XX

CC The invention relates to a new microarray which comprises probes for  
CC genes involved in psychoneuroendocrinimmune (PNI) activity. The  
CC microarray is useful in diagnosing a condition associated with PNI  
CC activity, such as CFS, type-2 diabetes, allergic condition, inflammation,  
CC cancer and infection. The present sequence represents a  
CC psychoneuroendocrinimmune gene expressed sequence tag. Note the  
CC specificatio mentions SEQ ID NO of up to 3314 but only sequences up to  
CC SEQ ID NO 1829 are provided.

XX

SQ Sequence 2907 BP; 820 A; 662 C; 759 G; 666 T; 0 U; 0 Other;

Query Match 83.2%; Score 15.8; DB 14; Length 2907;  
Best Local Similarity 89.5%; Pred. No. 7.7e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCCAATTGA 19

|||||

Db 733 CCCAATGCATCCCAACTTGA 751

Search completed: November 20, 2005, 17:43:50

Job time : 310.691 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2005, 16:52:01 ; Search time 2841.56 Seconds  
(without alignments)  
312.841 Million cell updates/sec

Title: US-10-627-757-27

Perfect score: 19

Sequence: 1 cccagtcgcatcccaattga 19

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_gsa1:\*  
9: gb\_gsa2:\*  
10: gb\_gsa3:\*  
11: gb\_gsa4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	19	100.0	389	2	BF358253 PM1-EN000
C 2	19	100.0	528	2	BF358268 PM1-EN000
C 3	19	100.0	606	6	CB216116 NISC nq01
C 4	18	94.7	478	1	AV695785 AV695785
C 5	17.4	91.6	449	9	AQ246894 HS 2181_B
C 6	17.4	91.6	644	10	CW185212 104_602_1
C 7	17.4	91.6	646	10	CW185213 104_602_1
C 8	17.4	91.6	655	11	CR879414 Sus scrofa
C 9	17.4	91.6	764	10	CW352728 fbb001f0
C 10	17.4	91.6	786	10	CW352729 fbb001f0
C 11	17.4	91.6	874	7	CK412531 AUF IgG11
C 12	17	89.5	435	8	W01386 YY67f11.r1
C 13	17	89.5	542	9	AZ145178 SP_0024_B
C 14	17	89.5	575	5	BX096335 BX096335
C 15	17	89.5	668	7	CN285965 170005321
C 16	17	89.5	681	10	AG176731 Pan trogl
C 17	17	89.5	698	6	CA173627 SCCST100
C 18	17	89.5	705	3	EM008141 603618031
C 19	17	89.5	720	8	DN381398 LIB38534
C 20	17	89.5	808	6	CB601006 AGENCOURT
C 21	16.4	86.3	237	8	H92995 Yv07e04.s1
C 22	16.4	86.3	265	5	BU782452 1179e07.x

C 23	16.4	86.3	303	8	N73178
C 24	16.4	86.3	312	10	CL703167
C 25	16.4	86.3	332	7	CK685160
C 26	16.4	86.3	334	1	AV690308
C 27	16.4	86.3	335	1	AV654794
C 28	16.4	86.3	347	1	AV660416
C 29	16.4	86.3	347	2	BI041584 PMO-NT031
C 30	16.4	86.3	359	1	AV651425
C 31	16.4	86.3	360	1	AA186919
C 32	16.4	86.3	362	1	H95586
C 33	16.4	86.3	364	8	T56720
C 34	16.4	86.3	371	1	AV655545
C 35	16.4	86.3	376	1	AV649876
C 36	16.4	86.3	383	10	CL698283
C 37	16.4	86.3	386	1	AV649842
C 38	16.4	86.3	390	1	AV659515
C 39	16.4	86.3	400	7	CN275794
C 40	16.4	86.3	400	10	CL700771
C 41	16.4	86.3	404	8	W24996
C 42	16.4	86.3	406	8	W24013
C 43	16.4	86.3	412	8	H73194
C 44	16.4	86.3	426	1	AA186416
C 45	16.4	86.3	433	6	CB067264

ALIGNMENTS

RESULT 1  
BF358253/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
PUBMED  
COMMENT  
FEATURES  
source

BF358253  
PM1-EN0007-130700-003-c10 EN0007 Homo sapiens cDNA, mRNA sequence.  
BF358253  
BF358253.1 GI:11317325  
EST.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
1 (bases 1 to 389)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,  
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
10737800  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?c1=PM1&t2=PM1-EN0007-  
130700-003-c10&t3=2000-07-13&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 388.  
Location/Qualifiers  
1..389  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="EN0007"  
/note="Organ: lung\_normal; Vector: puc18; Site\_1: Sma1;

Site2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

## ORIGIN

```
Query Match      100.0%; Score 19; DB 2; Length 389;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 19; Conservative 0; Mismatches 0; Indels
```

Qy 1 CCCAGTGCATCCAAATTGA 19  
|||  
pb 163 CCCAGTGCATCCAAATTGA 145

RESULT 2	BF358268/c	LOCUS	BF358268	528 bp	mRNA	linear	EST 22-NOV-2000
		DEFINITION	PM1-EN0007-240800-005-e05	EN0007	Homo sapiens	cDNA, mRNA	sequence.

ORGANISM	SOURCE	Ref.
Homo sapiens (human)		101.
Homo sapiens		
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.		

```

/clone_libs="EN0007"
/note=Organ: lung_normal; Vector: puc18; Site_1: SmaI;
site 2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. letters patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

```

## ORIGIN

Query Match 100.0%; Score 19; DB 2; Length 528;  
Best Local Similarity 100.0%; Pred. No. 76;

	Matches	19;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	1	CCCAGTGCATCCAAATTGA	19							
Dd	187	CCCAGTGCATCCAAATTGA	169							

RESULT	3
CB2161116	
LOCUS	
DEFINITION	CB216116 606 bp mRNA linear EST 06-FEB-2003 NISC nq01e05.y1 NTCHD HS_Ut2 Homo sapiens cDNA clone IMAGE:5937872 5'...mRNA sequence.

ORGANISM      Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

```

FEATURES
source
Location/Qualifiers
1. .606
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5937872"
/sex="female"
/tissue_type="normal endometrium, mid-secretory phase,
cycle day 23"
/lab_host="DH10B (TI-resistant)"
/clone_lib="NICHD HS Ut-2"
/note="Organ: uterus; Vector: pCMV-SPORT6.1.cdbb (Resgen,
Invitrogen Corporation); Site_1: NciI; Site_2: EcoRV;
cloned unidirectionally from microquantity amounts of mRNA
from normal endometrial tissue (mid-secretory phase, cycle
day 23). Average insert size 1.6 kb. Library constructed
bv Resgen (Invitrogen Corporation)."
```

## ORIGIN

Query Match	100.0%	Score 19;	DB 6;	Length 606;
Best Local Similarity	100.0%;	Pred. No. 77;		
Matches 19:	Conservative	0;	Mismatches	0;
	Indels	0;	Gaps	0;

Qy 1 CCCAGTGCATCCAAATTGA 19  
|||  
Db 351 CCCAGTGCATCCAAATTGA 369

RESULT 4	AV695785/c	AV695785	478 bp	mrna	linear	EST 16-JAN-2002
LOCUS	AV695785/c					
DEFINITION	AV695785	GKC Homo sapiens	cdna clone	GKCDZH03	5',	mrna sequence.
ACCESSION	AV695785					
VERSION	AV695785.1	GI:10297648				
KEYWORDS	EST.					
SOURCE	Homo sapiens	(human)				
ORGANISM	Homo sapiens					



Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

## AUTHORS

1 (bases 1 to 478)  
 Xu, X., Huang, J., Xu, Z., Qian, B., Zhu, Z., Yan, Q., Cai, T., Zhang, X., Xiao, H., Qu, J., Liu, F., Huang, Q., Cheng, Z., Li, N., Du, J., Hu, W., Shen, K., Lu, G., Fu, G., Zhong, M., Xu, S., Gu, W., Huang, W., Zhao, X., Hu, G., Gu, J., Chen, Z. and Han, Z.

**TITLE**  
 Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver

**JOURNAL**  
**PUBLISHED**  
**COMMENT**  
 Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)  
 11752456

Contact: Zeguang Han  
 Chinese National Human Genome Center at Shanghai  
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China  
 Tel: 86-21-50801919 (ex.45)  
 Fax: 86-21-50801922  
 Email: hanzg@hgc.sh.cn

This clone is available at CHGC in Shanghai.

## FEATURES

## source

1..478  
 Location/Qualifiers

/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="GKCDZH03"  
/tissue\_type="hepatocellular carcinoma"  
/dev\_stage="Adult"  
/lab\_host="SOLR"  
/clone\_lib="GKC"  
/note="Vector: pBluescript sk(-); Site\_1: EcoRI; Site\_2: XhoI"

## ORIGIN

Query Match 94.7%; Score 18; DB 1; Length 478;  
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CCAGTGCATCCCAATTGA 19

Db 215 CCAGTGCATCCCAATTGA 198

## RESULT 5

## AQ246894

## LOCUS

AQ246894 449 bp DNA linear GSS 06-OCT-1998  
 HS\_2181\_B1\_A09\_MF CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2181 Col=17 Row=B, genomic survey sequence.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

AQ246894.1 GI:3697076  
 Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

## AUTHORS

1 (bases 1 to 449)  
 Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and Hood, L.

**TITLE**  
 Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

## JOURNAL

## PUBLISHED

## COMMENT

Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
 10449764

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Sequence Tagged Connector  
 Plate: 2181 row: B column: 17  
 Class: BAC ends  
 High quality sequence stop: 449.

## FEATURES

## source

1..449  
 Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="Plate=2181 Col=17 Row=B"  
/sex="male"  
/clone\_lib="CIT Approved Human Genomic Sperm Library D"  
/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in E-Coli DH10B"

## ORIGIN

Query Match 91.6%; Score 17.4; DB 9; Length 449;  
 Best Local Similarity 94.7%; Pred. No. 5.2e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCCAATTGA 19

Db 184 CCCAGTTCATCCCAATTGA 202

## RESULT 6

## CW185212

## LOCUS

CW185212 644 bp DNA linear GSS 30-OCT-2004  
 104\_602\_11165944\_116\_36696\_052 Sorghum methylation filtered library (LibID: 104) Sorghum bicolor genomic clone 11165944, genomic survey sequence.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

CW185212 GI:54893540  
 Sorghum bicolor (sorghum)  
 Sorghum bicolor  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.

## REFERENCE

## AUTHORS

1 (bases 1 to 644)  
 Bedell, J.A., Budiman, M.A., Nunberg, A., Citek, R.W., Robbins, D., Jones, J., Flick, E., Rohlfing, T., Fries, J., Bradford, K., McMenamy, J., Smith, M., Holean, H., Roe, B.A., Wiley, G., Korf, I.F., Rabinowicz, P.D., Lakey, N., McCombie, W.R., Jeddalo, J.A. and Martienssen, R.A.

Sorghum genome sequencing by methylation filtration

PLoS Biol. 3 (1), e13 (2005)

## JOURNAL

## PUBLISHED

## COMMENT

15660154  
 Orion Genomics, LLC  
 Contact: Bedell JA  
 4041 Forest Park Ave, St. Louis, MO 63108, USA

Tel: 314 615 6979

Fax: 314 615 5975

Email: jbedell@oriongenomics.com

Seq primer: T3 Reverse

Plate: 602 row: n column: 16

Class: methylation filtered

High quality sequence stop: 644.

## FEATURES

## source

1..644  
 Location/Qualifiers  
/organism="Sorghum bicolor"  
/mol\_type="genomic DNA"  
/cultivar="ATx623"  
/db\_xref="taxon:4558"  
/clone="11165944"  
/clone\_lib="Sorghum methylation filtered library (LibID: 104)"

/note="Organ: leaf; Vector: pBCK(-); Site 1: HincII; DNA prepared from purified nuclei was randomly sheared, end-repaired, size fractionated to enrich for the 0.5 to 5 kb fraction, ligated into HincII-digested pBCK(-) vector and electroporated into E. coli cells. This is a methylation filtered library."

```

ORIGIN
Query Match          91.6%; Score 17.4; DB 10; Length 644;
Best Local Similarity 94.7%; Pred. No. 5.4e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCGAGTGCATCCAAATTGA 19
||| ||||| ||||| ||||| |||||
DB 195 CCGAGTGCATCCAAATTGA 213

RESULT 7
CW185213/c
LOCUS
DEFINITION
104 602 11165944 148 36695.052 Sorghum methylation filtered library
(LiBiD: 104) Sorghum bicolor genomic clone 11165944, genomic survey
sequence.
ACCESSION
CW185213
VERSION
GSS.
SOURCE
Sorghum bicolor (sorghum)
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
Clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 646)
Bedell,J.A., Budiman,M.A., Nunberg,A., Citek,R.W., Robbins,D.,
Jones,J., Flick,E., Rohlfing,T., Fries,J., Bradford,K.,
McMenamy,J., Smith,M., Holean,H., Roe,B.A., Wiley,G., Korf,I.F.,
Rabinowicz,P.D., Lakey,N., McCombie,W.R., Jeddelloh,J.A. and
Martienssen,R.A.
Sorghum genome sequencing by methylation filtration
PLOS Biol. 3 (1), e13 (2005)
15660154
Contact: Bedell JA
Orion Genomics, LLC
4041 Forest Park Ave, St. Louis, MO 63108, USA
Tel: 314 615 6979
Fax: 314 615 5975
Email: jbedell@oriongenomics.com
Plate: 602 row: n column: 16
Seq primer: SWfor Forward
Class: methylation filtered
High quality sequence stop: 646.
Location/Qualifiers
1 .646
/organism="Sorghum bicolor"
/mol_type="genomic DNA"
/cultivar="Atx623"
/db_xref="taxon:4558"
/clone="11165944"
/clone_lib="Sorghum methylation filtered library (LiBiD:
104)"_
/note="Organ: leaf; Vector: pBCKS(-); Site 1: HincII; DNA
prepared from purified nuclei was randomly sheared,
end-repaired, size fractionated to enrich for the 0.5 to 5
kb fraction, ligated into HincII-digested pBCKS(-) vector
and electroporated into E. coli cells. This is a
methylation filtered library."

ORIGIN
Query Match          91.6%; Score 17.4; DB 10; Length 646;
Best Local Similarity 94.7%; Pred. No. 5.4e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCGAGTGCATCCAAATTGA 19
||| ||||| ||||| ||||| |||||
DB 533 CCGAGTGCATCCAAATTGA 515

RESULT 8
CR879414/c
LOCUS
DEFINITION
Sus scrofa BES, genomic survey sequence.
ACCESSION
CR879414
VERSION
GSS.
KEYWORDS
GSS; Bac-end sequence BES; Genome Survey Sequence.
SOURCE
Sus scrofa (pig)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;
Sus.
1 (bases 1 to 655)
Rogel-Gaillard,C., Bourgeaux,N., Billault,A., Vaiman,M. and
Chardon,P.
Construction of a swine BAC library: application to the
characterization and mapping of porcine type C endoviral elements
Cytogetnet. Cell Genet. 85 (3-4), 205-211 (1999)
10449899
2 (bases 1 to 655)
Chardon,P., Iannuccelli,N., Roig,A., Dossat,C., Demars,J.,
Rogel-Gaillard,C., Roy,A., Schibler,L. and Milan,D.
A physical map of the swine genome
Unpublished
3 (bases 1 to 655)
Genoscope.
Direct Submission
Submitted (18-NOV-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr
- Web : www.genoscope.cns.fr)
Location/Qualifiers
1 .655
/organism="Sus scrofa"
/mol_type="genomic DNA"
/strain="Large White"
/db_xref="taxon:9823"
/clone="b10890G06"
/sex="male"
/cell_type="fibroblast"
/clone_lib="SBAB"
/note="Genoscope sequence ID : IH0AAA87AA07RM1"

FEATURES
source
Query Match          91.6%; Score 17.4; DB 11; Length 655;
Best Local Similarity 94.7%; Pred. No. 5.4e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCGAGTGCATCCAAATTGA 19
||| ||||| ||||| ||||| |||||
DB 53 CCGAGTGCATCCAAAGTGA 35

RESULT 9
CW352728/c
LOCUS
DEFINITION
CW352728 fbb001f015c10f0 Sorghum methylation filtered library (LiBiD: 104)
Sorghum bicolor genomic clone fbb001f015c10, genomic survey
sequence.
ACCESSION
CW352728
VERSION
GSS.
SOURCE
Sorghum bicolor (sorghum)
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
Clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 764)
Bedell,J.A., Budiman,M.A., Nunberg,A., Citek,R.W., Robbins,D.,
Jones,J., Flick,E., Rohlfing,T., Fries,J., Bradford,K.,
McMenamy,J., Smith,M., Holean,H., Roe,B.A., Wiley,G., Korf,I.F.,
Rabinowicz,P.D., Lakey,N., McCombie,W.R., Jeddelloh,J.A. and
Martienssen,R.A.
Sorghum genome sequencing by methylation filtration
PLOS Biol. 3 (1), e13 (2005)
15660154
Contact: Bedell JA

```

```

/db_xref="taxon:4538"
/clone="fab001f015c10"
/clone_lib="Sorghum methylation filtered library (LibID:
104)"
notes="Organ: leaf; Vector: pBCSK(-); Site 1: HincII; DNA
prepared from purified nuclei was randomly sheared,
end-repaired, size fractionated to enrich for the 0.5 to 5
kb fraction, ligated into HincII-digested pBCSK(-) vector
and electroporated into E. coli cells. This is a
methylation filtered library."

```

**source**

## ORIGIN

```

Query Match      91.5%; Score 17.4; DB 10; Length 786;
Best Local Similarity 94.7%; Pred. No. 5.6e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CCCAGTGCATCCAAATTGA 19
          |||||
Db      554 CCGAGTGCATCCAAATTGA 572

```

SOURCE	ORGANISM
ictalurius punctatus (channel callish)	ictalurius punctatus
	ictalurius punctatus

## 1729

**TITLE** 30,000 new catfish ESTs: new resources for functional analysis of genes involved in aquaculture performance traits

**JOURNAL** Unpublished (2004)

COCKATOO  
COCKATOO (2004)  
Contact: Liu ZJ  
COMMENT

and Molecular Biosciences  
Auburn University  
203 Swingle Hall, Auburn University, Auburn, AL 36849, USA  
Tel: 334 844 4054  
Fax: 334 844 9208  
Email: zliu@acesag.auburn.edu  
Seq primer: T7.

## THORS

```

FEATURES
source
1 874
Location/Qualifiers
/organism="Ictalurus punctatus"
/mol_type="mRNA"
/db_xref="taxon.7998"
/clone_lib="Ictalurus punctatus gill cDNA library"
/note="Organ: Gill; Vector: pSport1; Site_1: NotI; Site_2: SalI"

```

```

/mot_type="unqua-
/db_xref="taxon:7998"
/clone_lib=Ictalurus punctatus gill cDNA library"
/note="Organ: Gill; Vector: pSport1; Site_1: NotI; Site_2:
SalI"

```

Query Match	91.6%	Score 17.4;	DB 7;	Length 874;
Best Local Similarity	94.7%	Pred. No. 5.7e+02;		
Matches	18;	Conservative	0;	Mismatches 1;
				Indels 0;
Gaps	0;			
QY	1	CCCAAGTGCATCCAAATTGA	19	
Db	670	CCCTGTGTCATCCAAATTGA	652	

	RESULT 12				
	W01386				
	LORTIS				
	W01386	435 bp	mpna	linear	EST 18-APR-1996

```

DEFINITION      YV67f11.r1 Soares multiple sclerosis 2NbHMP Homo sapiens cDNA
                  clone IMAGE:278637 5', mRNA sequence.
ACCESSION       W01386
VERSION         W01386.1  GI:1273366
SOURCE         EST.
SOURCE         Homo sapiens (human)
ORGANISM        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
                  Homnidae; Homo.
REFERENCE       1 (bases 1 to 435)
AUTHORS         Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
                  Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
                  Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
                  Trevisakis,E., Waterson,K., Williamson,A., Wohlmann,P. and
                  Wilson,R.
TITLE           The WashU-Merck EST Project
JOURNAL         Unpublished (1995)
COMMENT         Contact: Wilson RK
                  Washington University School of Medicine
                  4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
                  Tel: 314 286 1800
                  Fax: 314 286 1810
                  Email: est@watson.wustl.edu
                  This clone is available royalty-free through LNL; contact the
                  IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES        Seq primer: ETPimer
                  High quality sequence stop: 392.
FEATURES        Location/Qualifiers
                  1..435
                   /organism="Homo sapiens"
                   /mol_type="mRNA"
                   /db_xref="GDB:3897013"
                   /db_xref="taxon:9606"
                   /clone="IMAGE:278637"
                   /sex="male"
                   /tissue_type="multiple sclerosis lesions"
                   /dev_stage="Age 46"
                   /lab_host="DH10B (ampicillin resistant)"
                   /clone_lib="Soares multiple sclerosis 2NbHMP"
                   /notes="Vector: pT7T3D (Pharmacia) with a modified
                   polylinker V TYPE: phagemid; Site 1: Not I; Site 2: Eco
                   RI; 1st strand cDNA was primed with a Not I - oligo(dT)
                   primer [5'
                   TGTTACCAATCTGAAGTGGAGCGCGCATTTTTTTTTTTTTTTT 3'],
                   double-stranded cDNA was size selected, ligated to Eco RI
                   adapters (Pharmacia), digested with Not I and cloned into
                   the Not I and Eco RI sites of a modified pT7T3 vector
                   (Pharmacia). Library went through one round of
                   normalization to a Cot = 5. Library constructed by Bento
                   Soares and M.Fatima Bonaldo. RNA from 4 multiple sclerosis
                   lesions from one patient was kindly provided by Dr. Kevin
                   G. Becker (NINDS/NIH)."
ORIGIN
Query Match      89.5%; Score 17; DB 8; Length 435;
Best Local Similarity 100.0%; Pred. No. 8.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATT 17
    |||||
Db 142 CCCAGTGCATCCAAATT 158

RESULT 13
AZ145178
LOCUS           AZ145178             542 bp      DNA      linear      GSS 28-AUG-2000
DEFINITION     SP_0024_B1_D04_T7 Strongylocentrotus purpuratus, purple sea urchin,
                  sperm genomic BAC library Strongylocentrotus purpuratus genomic
                  clone Plate=24 Col=7 Row=H, genomic survey sequence.
ACCESSION      AZ145178
VERSION        AZ145178.1  GI:8297081
KEYWORDS       GSS.

DEFINITION     YV67f11.r1 Soares multiple sclerosis 2NbHMP Homo sapiens cDNA
                  clone IMAGE:278637 5', mRNA sequence.
ACCESSION       W01386
VERSION         W01386.1  GI:1273366
SOURCE         EST.
SOURCE         Homo sapiens (human)
ORGANISM        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
                  Homnidae; Homo.
REFERENCE       1 (bases 1 to 435)
AUTHORS         Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
                  Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
                  Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
                  Trevisakis,E., Waterson,K., Williamson,A., Wohlmann,P. and
                  Wilson,R.
TITLE           The WashU-Merck EST Project
JOURNAL         Unpublished (1995)
COMMENT         Contact: Wilson RK
                  Washington University School of Medicine
                  4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
                  Tel: 314 286 1800
                  Fax: 314 286 1810
                  Email: est@watson.wustl.edu
                  This clone is available royalty-free through LNL; contact the
                  IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES        Seq primer: ETPimer
                  High quality sequence stop: 392.
FEATURES        Location/Qualifiers
                  1..435
                   /organism="Homo sapiens"
                   /mol_type="mRNA"
                   /db_xref="GDB:3897013"
                   /db_xref="taxon:9606"
                   /clone="IMAGE:278637"
                   /sex="male"
                   /tissue_type="multiple sclerosis lesions"
                   /dev_stage="Age 46"
                   /lab_host="DH10B (ampicillin resistant)"
                   /clone_lib="Soares multiple sclerosis 2NbHMP"
                   /notes="Vector: pT7T3D (Pharmacia) with a modified
                   polylinker V TYPE: phagemid; Site 1: Not I; Site 2: Eco
                   RI; 1st strand cDNA was primed with a Not I - oligo(dT)
                   primer [5'
                   TGTTACCAATCTGAAGTGGAGCGCGCATTTTTTTTTTTTTTTT 3'],
                   double-stranded cDNA was size selected, ligated to Eco RI
                   adapters (Pharmacia), digested with Not I and cloned into
                   the Not I and Eco RI sites of a modified pT7T3 vector
                   (Pharmacia). Library went through one round of
                   normalization to a Cot = 5. Library constructed by Bento
                   Soares and M.Fatima Bonaldo. RNA from 4 multiple sclerosis
                   lesions from one patient was kindly provided by Dr. Kevin
                   G. Becker (NINDS/NIH)."
ORIGIN
Query Match      89.5%; Score 17; DB 8; Length 435;
Best Local Similarity 100.0%; Pred. No. 8.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATT 17
    |||||
Db 142 CCCAGTGCATCCAAATT 158

RESULT 13
AZ145178
LOCUS           AZ145178             542 bp      DNA      linear      GSS 28-AUG-2000
DEFINITION     SP_0024_B1_D04_T7 Strongylocentrotus purpuratus, purple sea urchin,
                  sperm genomic BAC library Strongylocentrotus purpuratus genomic
                  clone Plate=24 Col=7 Row=H, genomic survey sequence.
ACCESSION      AZ145178
VERSION        AZ145178.1  GI:8297081
KEYWORDS       GSS.

```

```

SOURCE          Strongylocentrotus purpuratus
ORGANISM        Strongylocentrotus purpuratus
                  Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
                  Echinoidea; Euechinoidea; Echinacea; Echinoida;
                  Strongylocentrotidae; Strongylocentrotus.
REFERENCE       1 (bases 1 to 542)
AUTHORS         Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R.,
                  Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T.,
                  Wray,G.A., Etensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H.
                  and Hood,L.
TITLE           A sea urchin genome project: Sequence scan, virtual map, and
                  additional resources
JOURNAL         Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)
PUBMED         10920135
COMMENT         Contact: Cameron, RA, Davidson, EH, Hood, L
                  Division of Biology 156-29
                  California Institute of Technology
                  Pasadena California 91125, USA
                  Tel: (626) 395-8421
                  Fax: (626) 793-3047
                  Email: acameron@caltech.edu
                  Plate: 24 row: H column: 7
                  Seq primer: T7
                  Class: BAC ends
                  High quality sequence stop: 542.
FEATURES        Location/Qualifiers
                  1..542
                   /organism="Strongylocentrotus purpuratus"
                   /mol_type="genomic DNA"
                   /db_xref="taxon:7668"
                   /clone="Plate=24 Col=7 Row=H"
                   /clone_lib="Strongylocentrotus purpuratus, purple sea
                   urchin, sperm genomic BAC library"
                   /note="Organ: sperm; Vector: BAC3.6; BAC Clones in E-Coli
                   DH10B"
ORIGIN
Query Match      89.5%; Score 17; DB 9; Length 542;
Best Local Similarity 100.0%; Pred. No. 8.6e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATT 17
    |||||
Db 467 CCCAGTGCATCCAAATT 483

RESULT 14
BX096335
LOCUS           BX096335             575 bp      mRNA      linear      EST 04-FEB-2003
DEFINITION     IMAG9998022615 ; IMAGE:278637, mRNA sequence.
ACCESSION      IMAG9998022615
VERSION        BX096335
KEYWORDS       EST.
SOURCE         Homo sapiens (human)
ORGANISM        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
                  Homnidae; Homo.
REFERENCE       1 (bases 1 to 575)
AUTHORS         Ebert,L., Heil,O., Hennig,S., Neubert,P., Partsch,E., Peters,M.,
                  Radelof,U., Schneider,D. and Korn,B.
TITLE           Human Unigeneset - RZPD3
JOURNAL         Unpublished (2003)
COMMENT         Contact: Ina Rolfs
                  RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
                  Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
                  RZPD; IMAG998023615
                  RZPDLib; I.M.A.G.E. cDNA Clone Collection;
                  Human Unigeneset - RZPD3 (RZPDLib No.972)
                  http://www.rzpd.de/CloneCards/cgi-
                  bin/showLib.pl.cgi/response?libNo=972 Contact: Ina Rolfs
                  RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH

```

Heubnerweg 6, D-14059 Berlin, Germany  
Tel: +49 30 32639 101  
Fax: +49 30 32639 111  
www.rzpd.de

This clone is available royalty-free from RZPD;  
contact RZPD (clone@rzpd.de) for further information. Seq primer:  
M13r, Primer sequence: TTTCACACAGGAACGCTATGAC.

## FEATURES

source

Location/Qualifiers  
1. .575  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:998022615 ; IMAGE:278637"  
/sex="male"

/tissue\_type="multiple sclerosis lesions"

/dev\_stage="Age 46"

/lab\_host="DH10B (ampicillin resistant)"

/clone\_lib="Soares multiple sclerosis 2NBMSP"

/notes="Vector: pT73D (Pharmacia) with a modified

polylinker V.TYPE: phagemid; Site 1: Not I; Site 2: Eco

RI; 1st strand cDNA was primed with a Not I - oligo(dt)

primer [5',

TGTTACCAATCTCAAGTGGGAGCGCGCATTTTTTTTTTTTTTTT 3'].

double-stranded cDNA was size selected, ligated to Eco RI

adapters (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of a modified pT73 vector

(Pharmacia). Library went through one round of

normalization to a Cot = 5. Library constructed by Bento

Soares and M.Fatima Bonaldo. RNA from 4 multiple sclerosis

lesions from one patient was kindly provided by Dr. Kevin

G. Becker (NINDS/NIH). "

## ORIGIN

Query Match 89.5%; Score 17; DB 5; Length 575;

Best Local Similarity 100.0%; Pred. No. 8.7e+02;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATT 17

|||||

Db 142 CCCAGTGCATCCAAATT 158

## RESULT 15

CN285965

LOCUS

17000532195675 GRN\_ES Homo sapiens cDNA 5', mRNA sequence. EST 16-MAY-2004

ACCESSION

CN285965

VERSION

CN285965.1 GI:47302379

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

REFERENCE

AUTHORS

Brandenberger, R., Wei, H., Zhang, S., Lei, S., Murage, J., Fisk, G. J.,

Li, Y., Xu, C., Fang, R., Guegler, K., Rao, M. S., Mandalam, R.,

Lebkowski, J. and Stanton, L. W.

TITLE

Transcriptome characterization elucidates signaling networks that

control human ES cell growth and differentiation

JOURNAL

PUBMED

15146197

COMMENT

Contact: Brandenberger R

Regenerative Medicine

Geron Corporation

230 Constitution Drive, Menlo Park, CA 94025, USA

Tel: 650 473 8658

Fax: 650 473 7760

Email: rbrandenberger@geron.com

Insert Length: 668 Std Error: 0.00.

Location/Qualifiers

1. .668

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/tissue\_type="embryonic stem cells, cell lines H1, H7, and

H9"

/clone\_lib="GRN\_ES"

/notes="oligo dt primed, full-length enriched cDNA library

from undifferentiated hES cell lines H1 (p32), H7 (p29),

and H9 (p26) maintained in feeder-free conditions"

## ORIGIN

Query Match 89.5%; Score 17; DB 7; Length 668;

Best Local Similarity 100.0%; Pred. No. 8.9e+02;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATT 17

|||||

Db 625 CCCAGTGCATCCAAATT 641

## RESULT 16

AG176731

LOCUS

AG176731

DEFINITION

Pan troglodytes DNA, clone: RP43-048A24.TJ, genomic survey

sequence.

ACCESSION

AG176731

VERSION

AG176731.1 GI:16706411

KEYWORDS

GSS.

SOURCE

Pan troglodytes (chimpanzee)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Pan.

REFERENCE

AUTHORS

Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,

Totoki, Y., Watanabe, H. and Sakaki, Y.

TITLE

BAC end sequences of Library RPCI-43

JOURNAL

Unpublished

REFERENCE

AUTHORS

Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,

Totoki, Y., Watanabe, H. and Sakaki, Y.

TITLE

Direct Submission

JOURNAL

Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical

and Chemical Research (RIKEN), Genomic Sciences Center (GSC);

1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

(E-mail: chipbes@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/.

Tel: 81-45-503-9111, Fax: 81-45-503-9170)

COMMENT

Clones are derived from the chimpanzee BAC library RPCI-43 This BAC

end was generated during the R&D process and may have higher chance

of clone tracking errors.

PRIMERS

Sequencing: TJ

LIBRARY

Vector : PBACe3.6

R.Site 1 : EcoRI

R.Site 2 : EcoRI.

Location/Qualifiers

1. .681

/organism="Pan troglodytes"

/mol\_type="genomic DNA"

/db\_xref="taxon:9598"

/clone="RP43-048A24.TJ"

/sex="male"

/cell\_type="lymphocytes"

/clone\_lib="RPCI-43 Chimpanzee Male BAC Library"

## ORIGIN

Query Match 89.5%; Score 17; DB 10; Length 681;

Best Local Similarity 100.0%; Pred. No. 8.9e+02;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 CAGTGCATCCAAATTGA 19

|||||

Db 633 CAGTGCATCCAAATTGA 649

```

RESULT 17
CAI173627/c
LOCUS
DEFINITION
  CAI173627 698 bp mRNA linear EST 24-SEP-2003
  SCCST1002D06.g S71 Saccharum officinarum cDNA clone SCCST1002D06
  5', mRNA sequence.
ACCESSION
  CAI173627.1 GI:35100446
VERSION
  EST.
KEYWORDS
  Saccharum officinarum
SOURCE
  Saccharum officinarum
ORGANISM
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
  clade; Panicoideae; Andropogoneae; Saccharum; Saccharum officinarum
  complex.
REFERENCE
  1 (bases 1 to 698)
  Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
  The libraries that made SUCEST
  Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
  Contact: Arruda P
  Centro de Biologia Molecular e Engenharia Genetica
  Universidade Estadual de Campinas
  Caixa Postal 6010, 13083-970, Campinas SP, Brazil
  Tel: 55 19 3788 1137
  Fax: 55 19 3788 1089
  Email: parruda@unicamp.br
  Clone distribution: clone distribution information can be found
  through the Brazilian Clone Collection Center (BCCC) at
  http://www.bcccenter.fcav.unesp.br
  Plate: 002 row: D column: 06
  Seq primer: T7 Promoter Primer.
  Location/Qualifiers
    1..698
      /organism="Saccharum officinarum"
      /mol_type="mRNA"
      /db_xref="taxon:4547"
      /clone="SCCST1002D06"
      /lab_host="DH10B"
      /clone_lib="ST1"
      /note="Organ: First apical stalk internodes of adult
      plants; Vector: pSport1; Site 1: SalI; Site 2: NotI; An
      unidirectional cDNA library generated from [first apical
      stalk internodes of adult plants]. cDNA was prepared from
      polyA+ mRNA using SuperScript Plasmid System Kit
      (Invitrogen). The double-strand cDNAs were fractionated
      in a sepharose CL-2B 40cm-columns and fragments sizing
      between 0.8 and 1.5 Kb were directionally cloned into the
      vector. Details of each source of RNA and library
      construction can be obtained at
      http://sucest.lad.ic.unicamp.br/public"
FEATURES
  source
    1..698
      /organism="Saccharum officinarum"
      /mol_type="mRNA"
      /db_xref="taxon:4547"
      /clone="SCCST1002D06"
      /lab_host="DH10B"
      /clone_lib="ST1"
      /note="Organ: First apical stalk internodes of adult
      plants; Vector: pSport1; Site 1: SalI; Site 2: NotI; An
      unidirectional cDNA library generated from [first apical
      stalk internodes of adult plants]. cDNA was prepared from
      polyA+ mRNA using SuperScript Plasmid System Kit
      (Invitrogen). The double-strand cDNAs were fractionated
      in a sepharose CL-2B 40cm-columns and fragments sizing
      between 0.8 and 1.5 Kb were directionally cloned into the
      vector. Details of each source of RNA and library
      construction can be obtained at
      http://sucest.lad.ic.unicamp.br/public"
ORIGIN
  Query Match 89.5%; Score 17; DB 6; Length 698;
  Best Local Similarity 100.0%; Pred. No. 8.9e+02;
  Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

  QY 1 CCCAGTGCATCCAAATT 17
  |||||
  DB 648 CCCAGTGCATCCAAATT 632

RESULT 18
BM008141/c
LOCUS
DEFINITION
  BM008141 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:5451078 5',
  mRNA sequence.
ACCESSION
  BM008141
VERSION
  EST.
KEYWORDS
  Homo sapiens (human)
SOURCE
  Homo sapiens
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

REFERENCE
  AUTHORS
  TITLE
  JOURNAL
  COMMENT
  1 (bases 1 to 705)
  NIH-MGC http://mgc.nci.nih.gov/.
  National Institutes of Health, Mammalian Gene Collection (MGC)
  Unpublished (1999)
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs-r@mail.nih.gov
  Tissue Procurement: Dr. Mark Watson
  cDNA Library Preparation: Ling Hong/Rubin Laboratory
  cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
  DNA Sequencing by: Incyte Genomics, Inc.
  Clone distribution: MGC clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at:
  http://image.llnl.gov
  Plate: LLCM1940 row: n column: 07
  High quality sequence stop: 658.
  Location/Qualifiers
    1..705
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="IMAGE:5451078"
      /lab_host="DH10B (phage-resistant)"
      /clone_lib="NIH_MGC_113"
      /note="Organ: spleen; Vector: pOTB7; Site 1: XhoI; Site 2:
      EcoRI; cDNA made by oligo-dT priming. Directionally cloned
      into EcoRI/XhoI sites using the following 5' adaptor:
      GGCACGAG(G). Library constructed by Ling Hong in the
      laboratory of Gerald M. Rubin (University of California,
      Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
      Superscript II RT (Life Technologies). Note: this is a
      NIH_MGC Library."
FEATURES
  source
    1..705
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="IMAGE:5451078"
      /lab_host="DH10B (phage-resistant)"
      /clone_lib="NIH_MGC_113"
      /note="Organ: spleen; Vector: pOTB7; Site 1: XhoI; Site 2:
      EcoRI; cDNA made by oligo-dT priming. Directionally cloned
      into EcoRI/XhoI sites using the following 5' adaptor:
      GGCACGAG(G). Library constructed by Ling Hong in the
      laboratory of Gerald M. Rubin (University of California,
      Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
      Superscript II RT (Life Technologies). Note: this is a
      NIH_MGC Library."
ORIGIN
  Query Match 89.5%; Score 17; DB 3; Length 705;
  Best Local Similarity 100.0%; Pred. No. 8.9e+02;
  Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

  QY 1 CCCAGTGCATCCAAATT 17
  |||||
  DB 496 CCCAGTGCATCCAAATT 480

RESULT 19
DN381398
LOCUS
DEFINITION
  DN381398 720 bp mRNA linear EST 07-MAR-2005
  LIB38534_052_A08.T7.1 LIB38534 Canis familiaris cDNA clone
  LIB38534_52_A08, mRNA sequence.
ACCESSION
  DN381398
VERSION
  EST.
KEYWORDS
  Canis familiaris (dog)
SOURCE
  Canis familiaris
ORGANISM
  Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
  Eukaryota; Metazoa; Chordata; Laurasiatheria; Carnivora; Canidae;
  Canis.
REFERENCE
  1 (bases 1 to 720)
  Statens,N.R.
  Direct Submission (Statens,N.R.)
  Unpublished (2005)
  Contact: Nick Statens
  Tel: 656 247 6855
  Email: nicholas.r.statens@pfizer.com.
  Location/Qualifiers
    1..720
      /organism="Canis familiaris"
      /mol_type="mRNA"
      /db_xref="taxon:9615"
      /clone="LIB38534_52_A08"
      /tissue_type="unknown"
      /lab_host="DH10B"
      /clone_lib="LIB38534"

```

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 705)  
NIH-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Dr. Mark Watson  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLCM1940 row: n column: 07  
High quality sequence stop: 658.

FEATURES  
source

1..705  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5451078"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_113"  
/note="Organ: spleen; Vector: pOTB7; Site 1: XhoI; Site 2:  
EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCACGAG(G). Library constructed by Ling Hong in the  
laboratory of Gerald M. Rubin (University of California,  
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
Superscript II RT (Life Technologies). Note: this is a  
NIH\_MGC Library."

ORIGIN

Query Match 89.5%; Score 17; DB 3; Length 705;  
Best Local Similarity 100.0%; Pred. No. 8.9e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATT 17  
|||||  
DB 496 CCCAGTGCATCCAAATT 480

RESULT 19

DN381398  
LOCUS  
DEFINITION  
DN381398 720 bp mRNA linear EST 07-MAR-2005  
LIB38534\_052\_A08.T7.1 LIB38534 Canis familiaris cDNA clone  
LIB38534\_52\_A08, mRNA sequence.  
ACCESSION  
DN381398  
VERSION  
EST.  
KEYWORDS  
Canis familiaris (dog)  
SOURCE  
Canis familiaris  
ORGANISM  
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;  
Eukaryota; Metazoa; Chordata; Laurasiatheria; Carnivora; Canidae;  
Canis.

REFERENCE

1 (bases 1 to 720)  
Statens,N.R.  
Direct Submission (Statens,N.R.)  
Unpublished (2005)  
Contact: Nick Statens  
Tel: 656 247 6855  
Email: nicholas.r.statens@pfizer.com.

FEATURES

source  
1..720  
/organism="Canis familiaris"  
/mol\_type="mRNA"  
/db\_xref="taxon:9615"  
/clone="LIB38534\_52\_A08"  
/tissue\_type="unknown"  
/lab\_host="DH10B"  
/clone\_lib="LIB38534"

```
ORIGIN
Query Match      89.5%; Score 17; DB 8; Length 720;
Best Local Similarity 100.0%; Pred. No. 9e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 CAGTGCATCCAAATTGA 19
|||||
Db 359 CAGTGCATCCAAATTGA 375

RESULT 20
CB601006/c
LOCUS
DEFINITION
AGENCOURT 13034890 NIH MGC 176 Mus musculus cDNA clone
IMAGE:30312514 5', mRNA sequence.
ACCESSION
CB601006
VERSION
CB601006.1 GI:29518862
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidae; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 808)
AUTHORS
NIH-MGC http://mgc.nci.nih.gov/.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cga@bbsfemail.nih.gov
Tissue Procurement: Dr. Michael Brownstein
cDNA Library Preparation: Michael Brownstein Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDCM97 row: j column: 11
High quality sequence stop: 502.
FEATURES
Location/Qualifiers
1..808
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30312514"
/lab_host="DH10B (TI-phage-resistant)"
/clone_lib="NIH_MGC_176"
/notes="Organ: kidney; Vector: pDNR-LIB; Site_1: SfiI
(ggccattggcc); Site_2: SfiI (ggccctcgcc); cDNA made
by oligo-dT priming and directionally cloned. 5' and 3'
adaptors were used in cloning as follows:
5'-AACCATGTGTATCAAGCAGCGATGGCATACGCGCGG-3' and
5'-ATTCTAGAGCCGAGCGCGGACATG-dt(30)NN-3'. Full-length
enriched library was constructed using the Clontech
Creator SMART kit and size-selected to contain the 0.5 kb
size fraction. Library created in the laboratory of M.
Brownstein (NIMH, NIH). Note: this is a NIH_MGC Library."
ORIGIN
Query Match      89.5%; Score 17; DB 6; Length 808;
Best Local Similarity 100.0%; Pred. No. 9.1e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATT 17
|||||
Db 642 CCCAGTGCATCCAAATT 626

RESULT 21
H92995/c
LOCUS
DEFINITION
yv07e04.s1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
IMAGE:242046 3' similar to gb:X56411_rnal ALCOHOL DEHYDROGENASE
CLASS II PI CHAIN (HUMAN);, mRNA sequence.
ACCESSION
H92995
VERSION
H92995.1 GI:1099323
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE
1 (bases 1 to 237)
AUTHORS
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lemon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevasakis,E., Waterston,R., Williamson,A., Wohldmann,P. and
Wilson,R.
TITLE
The WashU-Merck EST Project
JOURNAL
Unpublished (1995)
COMMENT
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence starts: 1
High quality sequence stops: 1
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Trace considered overall poor quality
Seq primer: ml3 -40 primer
High quality sequence stop: 1.
FEATURES
Location/Qualifiers
1..237
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:3791179"
/db_xref="taxon:9606"
/clone="IMAGE:242046"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares fetal liver spleen INFLS"
/notes="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
with a modified polylinker; Site 1: Pac I; Site 2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5' AACTGGAGATTAATTAAGATCTTTTCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Patima Bonaldo."
ORIGIN
Query Match      86.3%; Score 16.4; DB 8; Length 237;
Best Local Similarity 94.4%; Pred. No. 1.6e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCAGTGCATCCAAATTGA 19
|||||
Db 74 CCAGTGCATCCAGATTGA 57

RESULT 22
BU782452
LOCUS
DEFINITION
BU782452.1 GI:23825721
sequence.
ACCESSION
BU782452
VERSION
BU782452.1
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
```

ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Muridae; Murinae; Mus.	
AUTHORS	1 (bases 1 to 265)	
	Melton,D., Brown,J., Kenty,G., Pezmutt,A., Lee,C., Kaestner,K., Lemishka,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S., Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A., Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas,M., Gibbons,M., McCann,R., Cole,R., Teagareishvili,R., Williams,T., Jackson,Y. and Bowers,Y.	
TITLE	Endocrine Pancreas Consortium	
JOURNAL	Unpublished (2000)	
COMMENT	Other_ESTs: i179e07.y1 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue Endocrine Pancreas Consortium Harvard University, Howard Hughes Medical Institute Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138 Tel: 617-495-1812 Fax: 617-495-8557 Email: dmelton@biohp.harvard.edu Pancreas was obtained from Gerard Gradwohl (PNAS 97 P1607-1611, 2000) Library was constructed by Catherine Lee DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Dr. Marie Searce (msearce@mail.med.upenn.edu) Trace considered overall poor quality Seq primer: -40UP from Gibco High quality sequence stop: 1.	
FEATURES	Location/Qualifiers	
source	1..265	
	/organism="Mus musculus"	
	/mol_type="mRNA"	
	/strain="129/Sv x CD1"	
	/db_xref="taxon:10090"	
	/clone="IMAGE:"	
	/dev_stage="p.c. 14.5"	
	/lab_host="E. coli-DH12S (GIBCO)"	
	/clone_lib="Kaestner ng3 -"	
	/note="Organ: pancreas; Vector: pSPOR2 (GIBCO); Site 1: Not 1; Site 2: Sal I; The library was prepared by Catherine S' Lee and has not been published. The pancreas was obtained from Gerard Gradwohl (PNAS 97 P1607-1611, 2000). The cDNA's were prepared with an oligo containing a NotI site, and SalI linkers were added to the ends. The inserts were cut with NotI before being cloned into the NotI-SalI sites in the vectors. This is one of two libraries, ng3 wt and ng3 -/- . The ng3 -/- library is in pSPOR2, T7 promoter is 3'."	
ORIGIN		
Query Match	86.3%; Score 16.4; DB 5; Length 265;	
Best Local Similarity	94.4%; Pred. No. 1.6e+03;	
Matches	17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
OY	1 CCAGTGCATCCAAATTG 18	
Db	219 CCAGTTCATCCAAATTG 236	
RESULT 23		
N73178/c		
LOCUS	N73178 303 bp mRNA linear EST 28-JAN-1997	
DEFINITION	Yv45g10.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone IMAGE:245730 5' similar to gb:X56411.rnal ALCOHOL DEHYDROGENASE CLASS II PI CHAIN (HUMAN); mRNA sequence.	
ACCESSION	N73178	
VERSION	N73178.1 GI:1230282	
KEYWORDS	EST.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
REFERENCE	1 (bases 1 to 303)	
AUTHORS	Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B., Chisoso,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le.M., Le.N., Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevasis,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.	
TITLE	Generation and analysis of 280,000 human expressed sequence tags	
JOURNAL	Genome Res. 6 (9), 807-828 (1996)	
COMMENT	8889549 Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available royalty-free through LBNL; contact the IMAGE Consortium (info@image.lbnl.gov) for further information. Insert Length: 1037 Std Error: 0.00 Seq primer: reverse Et High quality sequence stop: 176.	
FEATURES	Location/Qualifiers	
source	1..303	
	/organism="Homo sapiens"	
	/mol_type="mRNA"	
	/db_xref="GDB:3794976"	
	/db_xref="taxon:9606"	
	/clone="IMAGE:245730"	
	/sex="male"	
	/dev_stage="20 week-post conception fetus"	
	/lab_host="DH10B (ampicillin resistant)"	
	/clone_lib="Soares fetal liver spleen INFLS"	
	/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site 1: Pac I; Site 2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer 15' AACTGGAAGATTAATTAAGATCTTTTCTTTTCTTTTCTTTT 3', double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."	
ORIGIN		
Query Match	86.3%; Score 16.4; DB 8; Length 303;	
Best Local Similarity	94.4%; Pred. No. 1.7e+03;	
Matches	17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
OY	2 CCAGTGCATCCAAATTGA 19	
Db	73 CCAGTGCATCCAGATTGA 56	
RESULT 24		
CL703167		
LOCUS	SP_Ba0093J17.f SP_Ba Sorghum propinquum genomic clone	
DEFINITION	SP_Ba0093J17 5', genomic survey sequence.	
ACCESSION	CL703167	
VERSION	CL703167.2 GI:55988851	
KEYWORDS	GSS.	
SOURCE	Sorghum propinquum	
ORGANISM	Sorghum propinquum	
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.	
REFERENCE	1 (bases 1 to 312)	
AUTHORS	Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Kudrna,D., Muller,C., Soderlund,C., Bowers,J.E., Paterson,A.H. and Wing,R.	
TITLE	Physical mapping of the sorghum genome	
JOURNAL	Unpublished (2004)	



COMMENT On Nov 23, 2004 this sequence version replaced gi:50269442.

Contact: Rod A. Wing  
Arizona Genomics Institute  
University of Arizona  
Forbes Building Room 303, Tucson, AZ 85721-0036, USA  
Tel: 520 626 9595  
Fax: 520 621 1259  
Email: rwing@genome.arizona.edu

PCR Primers

FORWARD: TAA TAC GAC TCA CTA TAG GG

BACKWARD: CAC TCA TTA GGC ACC CCA

Plate: 0093 row: J column: 17

Seq primer: TAA TAC GAC TCA CTA TAG GG

Class: BAC ends.

Location/Qualifiers

FEATURES

source

1. .312

/organism="Sorghum propinquum"

/mol\_type="genomic DNA"

/db\_xref="taxon:132711"

/clone\_lib="SP\_Ba"

/notes="Vector: pBeloBAC11; Site 1: HindIII; Site 2: HindIII; Paterson lab BAC library (HindIII)"

ORIGIN

Query Match 86.3%; Score 16.4; DB 10; Length 312;

Best Local Similarity 94.4%; Pred. No. 1.7e+03;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCAGTGCATCCAAATTGA 19

|||||

Db 132 CGAGTGCATCCAAATTGA 149

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

/dev\_stages="7 Different embryonic Stages(Prom just

fertilized Embryos to 72 hours just hatched baby fish)"

/lab\_host="DH10B"

/clone\_lib="GISZF001\_ra"

/note="Vector: pDNR-LIB; Site 1: Sfi A (GCCATTACGGCC);

Site 2: Sfi B (GGCGAGGCGCC); Priming method: Sfi-(dT)30

Primed; Priming sequence:

5.ATTCTAGAGCGGAGCGCGGCACATG(T)30VN; Directionally

cloned, 5' cloning site: Sfi A site GCCATTACGGCC; 5'

linker/adaptor sequence: 5.AAGCAGTGTATCAACGAGAGTGGCC;

3' cloning site: Sfi B site GGCGAGGCGGCC; 3'

linker/adaptor sequence: same as the priming sequence;

Average insert size: 2kb; For PCR insert analysis: Use

M13 Forward and reverse primers; Library Amplified;

Recombinants (inserts): 98%; Library complexity: 5x10<sup>6</sup>;

Full-length construction (method): SMART, a Clontech

method The pooled tissue RNA was collected and used to

construct full length enriched cDNA library and also

served as template to synthesize complex first strand cDNA

probe. Two high density colony arrays were made from over

110K cDNA clones and hybridized with the probes. Low

intensity clones were selected as they represented rare

expressed clones. The hybridization intensities for all

clones span from 0 to 1.8 million counts and the low

abundant class ranged from 0 to 13,000."

ORIGIN

Query Match 86.3%; Score 16.4; DB 7; Length 332;

Best Local Similarity 94.4%; Pred. No. 1.7e+03;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATTG 18

|||||

Db 205 CCCAGTGCATCCAAATG 188

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

COMMENT On Nov 23, 2004 this sequence version replaced gi:50269442.

Contact: Rod A. Wing  
Arizona Genomics Institute  
University of Arizona  
Forbes Building Room 303, Tucson, AZ 85721-0036, USA  
Tel: 520 626 9595  
Fax: 520 621 1259  
Email: rwing@genome.arizona.edu

PCR Primers

FORWARD: TAA TAC GAC TCA CTA TAG GG

BACKWARD: CAC TCA TTA GGC ACC CCA

Plate: 0093 row: J column: 17

Seq primer: TAA TAC GAC TCA CTA TAG GG

Class: BAC ends.

Location/Qualifiers

FEATURES

source

1. .312

/organism="Sorghum propinquum"

/mol\_type="genomic DNA"

/db\_xref="taxon:132711"

/clone\_lib="SP\_Ba"

/notes="Vector: pBeloBAC11; Site 1: HindIII; Site 2: HindIII; Paterson lab BAC library (HindIII)"

ORIGIN

Query Match 86.3%; Score 16.4; DB 10; Length 312;

Best Local Similarity 94.4%; Pred. No. 1.7e+03;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCAGTGCATCCAAATTGA 19

|||||

Db 132 CGAGTGCATCCAAATTGA 149

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

COMMENT On Nov 23, 2004 this sequence version replaced gi:50269442.

Contact: Rod A. Wing  
Arizona Genomics Institute  
University of Arizona  
Forbes Building Room 303, Tucson, AZ 85721-0036, USA  
Tel: 520 626 9595  
Fax: 520 621 1259  
Email: rwing@genome.arizona.edu

PCR Primers

FORWARD: TAA TAC GAC TCA CTA TAG GG

BACKWARD: CAC TCA TTA GGC ACC CCA

Plate: 0093 row: J column: 17

Seq primer: TAA TAC GAC TCA CTA TAG GG

Class: BAC ends.

Location/Qualifiers

FEATURES

source

1. .312

/organism="Sorghum propinquum"

/mol\_type="genomic DNA"

/db\_xref="taxon:132711"

/clone\_lib="SP\_Ba"

/notes="Vector: pBeloBAC11; Site 1: HindIII; Site 2: HindIII; Paterson lab BAC library (HindIII)"

ORIGIN

Query Match 86.3%; Score 16.4; DB 10; Length 312;

Best Local Similarity 94.4%; Pred. No. 1.7e+03;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCAGTGCATCCAAATTGA 19

|||||

Db 132 CGAGTGCATCCAAATTGA 149

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

```

/dev_stage="Adult"
/lab_host="SOLR"
/clone_lib="GKC"
/notes="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

ORIGIN
Query Match      86.3%; Score 16.4; DB 1; Length 334;
Best Local Similarity 94.4%; Pred. No. 1.7e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCAGTGCATCCCAATTGA 19
|||||
Db 225 CCAGTGCATCCAGATTGA 208

RESULT 27
AV654794/c
LOCUS
DEFINITION AV654794 GLC Homo sapiens cDNA clone GLCDBE12 3', mRNA sequence.
ACCESSION AV654794
VERSION AV654794.1 GI:9875808
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 335)
Xiao,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
Hu,G., Gu,J., Chen,Z. and Han,Z.
Insight into hepatocellular carcinogenesis at transcriptome level
by comparing gene expression profiles of hepatocellular carcinoma
with those of corresponding noncancerous liver
Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
11752456
Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919 (ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
Location/Qualifiers
1. .335
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="GLCDBE12"
/tissue_type="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/clone_lib="GKC"
/notes="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

ORIGIN
Query Match      86.3%; Score 16.4; DB 1; Length 335;
Best Local Similarity 94.4%; Pred. No. 1.7e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCAGTGCATCCCAATTGA 19
|||||
Db 207 CCAGTGCATCCAGATTGA 190

RESULT 28
AV660416/c
LOCUS
DEFINITION AV660416 GLC Homo sapiens cDNA clone GLCDBE12 3', mRNA sequence.
ACCESSION AV660416
VERSION AV660416.1 GI:9881430
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 347)
Xiao,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
Hu,G., Gu,J., Chen,Z. and Han,Z.
Insight into hepatocellular carcinogenesis at transcriptome level
by comparing gene expression profiles of hepatocellular carcinoma
with those of corresponding noncancerous liver
Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
11752456
Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919 (ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
Location/Qualifiers
1. .347
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="GLCDBE12"
/tissue_type="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/clone_lib="GKC"
/notes="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

ORIGIN
Query Match      86.3%; Score 16.4; DB 1; Length 347;
Best Local Similarity 94.4%; Pred. No. 1.7e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCAGTGCATCCCAATTGA 19
|||||
Db 274 CCAGTGCATCCAGATTGA 257

RESULT 29
BI041584/c
LOCUS
DEFINITION PMO-NT0319-230301-001-f03 NT0319 Homo sapiens cDNA, mRNA sequence.
ACCESSION BI041584
VERSION BI041584.1 GI:14448210
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 347)
Dias Neto,B., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
10737800
Contact: Simpson A.J.G.

```

```

ACCESSION AV660416
VERSION AV660416.1 GI:9881430
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 347)
Xiao,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
Hu,G., Gu,J., Chen,Z. and Han,Z.
Insight into hepatocellular carcinogenesis at transcriptome level
by comparing gene expression profiles of hepatocellular carcinoma
with those of corresponding noncancerous liver
Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
11752456
Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919 (ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
Location/Qualifiers
1. .347
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="GLCDBE12"
/tissue_type="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/clone_lib="GKC"
/notes="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

ORIGIN
Query Match      86.3%; Score 16.4; DB 1; Length 347;
Best Local Similarity 94.4%; Pred. No. 1.7e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCAGTGCATCCCAATTGA 19
|||||
Db 274 CCAGTGCATCCAGATTGA 257

RESULT 29
BI041584/c
LOCUS
DEFINITION PMO-NT0319-230301-001-f03 NT0319 Homo sapiens cDNA, mRNA sequence.
ACCESSION BI041584
VERSION BI041584.1 GI:14448210
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 347)
Dias Neto,B., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
10737800
Contact: Simpson A.J.G.

```

Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?ti=PMO&t2=PMO-NT0319-  
230301-001-f03st3-2001-03-23&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 346.

#### FEATURES

Location/Qualifiers  
1..347  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="NT0319"

/notes="Organ: nervous tumor; Vector: puc18; Site 1: SmaI;  
Site 2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

#### ORIGIN

Query Match 86.3%; Score 16.4; DB 2; Length 347;  
Best Local Similarity 94.4%; Pred. No. 1.7e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATTG 18  
|||||  
Db 248 CCCAGTGCATCCAAATTG 231  
|||||

#### RESULT 30

AV651425/c 359 bp mRNA linear EST 15-JAN-2002  
LOCUS AV651425 GLC Homo sapiens cDNA clone GLCCPF10 3', mRNA sequence.  
DEFINITION AV651425  
ACCESSION AV651425  
VERSION AV651425.1 GI:9872439  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.

#### REFERENCE

1 (bases 1 to 359)  
Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,  
Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,  
Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,  
Hu,G., Gu,J., Chen,Z. and Han,Z.

Insight into hepatocellular carcinogenesis at transcriptome level  
by comparing gene expression profiles of hepatocellular carcinoma  
with those of corresponding noncancerous liver

Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)

#### JOURNAL

PUBLISHED 11732456

CONTACT: Zeguang Han

Chinese National Human Genome Center at Shanghai

351 Guo Shoujing Road, Shanghai Hi-Tech Park, Pudong, Shanghai

201203, P. R. China

Tel: 86-21-50801919 (ex.45)

Fax: 86-21-50801922

Email: hanzg@chgc.sh.cn

This clone is available at CHGC in Shanghai.

#### FEATURES

Location/Qualifiers  
1..359  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"

/clone="GLCCPF10"  
/tissue\_type="corresponding non cancerous liver tissue"  
/dev\_stage="Adult"  
/lab\_host="SOLR"  
/clone\_lib="GLC"  
/note="Vector: pBluescript sk(-); Site\_1: EcoRI; Site\_2:  
XhoI"

#### ORIGIN

Query Match 86.3%; Score 16.4; DB 1; Length 359;  
Best Local Similarity 94.4%; Pred. No. 1.7e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCAGTGCATCCAAATTGA 19  
|||||  
Db 109 CCAGTGCATCCAGATTGA 92  
|||||

#### RESULT 31

AA186919/c

LOCUS

DEFINITION AA186919 360 bp mRNA linear EST 10-MAR-1998  
zp66b05.r1 Stratagene endothelial cell 937223 Homo sapiens cDNA  
clone IMAGE:625137 5', similar to contains Alu repetitive  
element; contains element LIR9 repetitive element ;, mRNA sequence.

ACCESSION AA186919

VERSION AA186919.1 GI:1775046

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.

#### REFERENCE

1 (bases 1 to 360)

Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,

Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,

Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,

Theising,B., White,Y., Wyllie,T., Waterston,R. and Wilson,R.

WashU-NCI human EST Project

Unpublished (1997)

CONTACT: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.lnl.gov) for further information.

Insert Length: 1368 Std Error: 0.00

Seq primer: -28M13 rev2 from Amersham.

#### FEATURES

Location/Qualifiers  
1..360  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="GDB:5047291"  
/db\_xref="taxon:9606"  
/clone="IMAGE:625137"  
/dev\_stage="umbilical vein, 1 passage"  
/lab\_host="SOLR (kanamycin resistant)"  
/clone\_lib="Stratagene endothelial cell 937223"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
XhoI; Cloned unidirectionally. Primer: Oligo dt.

Umbilical vein endothelial cells, passaged once. Average

insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor

sequence: 5' GAATTCGGCAG 3' -3' adaptor sequence: 5'

CTCGAGTTTTTTTTTTTTTTT 3'

#### ORIGIN

Query Match 86.3%; Score 16.4; DB 1; Length 360;  
Best Local Similarity 94.4%; Pred. No. 1.7e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATTG 18  
|||||



**TITLE**  
Hu, G., Gu, J., Chen, Z., and Han, Z.  
Insight into hepatocellular carcinogenesis at transcriptome level  
by comparing gene expression profiles of hepatocellular carcinoma  
with those of corresponding noncancerous liver

**JOURNAL**  
Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)

**PUBMED**  
11752456

**COMMENT**  
On Aug 22, 2000 this sequence version replaced gi:9876559.  
Contact: Zeguang Han  
Chinese National Human Genome Center at Shanghai  
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai  
201203, P. R. China  
Tel: 86-21-50801919 (ex.45)  
Fax: 86-21-50801922  
Email: hanzg@chgc.sh.cn  
This clone is available at CHGC in Shanghai.

**FEATURES**

**source**  
1. .371  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="GLCEHH05"  
/tissue\_type="corresponding non cancerous liver tissue"  
/dev\_stage="Adult"  
/lab\_host="SOLR"  
/clone\_lib="GLC"  
/notes="Vector: pBluescript sk(-); Site\_1: EcoRI; Site\_2:  
XhoI"

**ORIGIN**

**Query Match** 86.3%; Score 16.4; DB 1; Length 371;  
**Best Local Similarity** 94.4%; Pred. No. 1.7e+03;  
**Matches** 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

**QY** 2 CCAGTGCATCCAAATTGA 19  
|||||  
**Db** 274 CCAGTGCATCCAGATTGA 257

**RESULT 35**  
**LOCUS** AV649876/c 376 bp mRNA linear EST 15-JAN-2002  
**DEFINITION** AV649876 GLC Homo sapiens cDNA clone GLCBYF06 3', mRNA sequence.  
**ACCESSION** AV649876  
**VERSION** AV649876.1 GI:9870890  
**KEYWORDS** EST.

**SOURCE** Homo sapiens (human)  
**ORGANISM** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

**REFERENCE** 1 (bases 1 to 376)  
**AUTHORS** Xu, X., Huang, J., Xu, Z., Qian, B., Zhu, Z., Yan, Q., Cai, T., Zhang, X.,  
Xiao, H., Qu, J., Liu, F., Huang, Q., Cheng, Z., Li, N., Du, J., Hu, W.,  
Shen, K., Lu, G., Fu, G., Zhong, M., Xu, S., Gu, W., Huang, W., Zhao, X.,  
Hu, G., Gu, J., Chen, Z. and Han, Z.

**TITLE**  
Insight into hepatocellular carcinogenesis at transcriptome level  
by comparing gene expression profiles of hepatocellular carcinoma  
with those of corresponding noncancerous liver

**JOURNAL**  
Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)

**PUBMED**  
11752456

**COMMENT**  
Contact: Zeguang Han  
Chinese National Human Genome Center at Shanghai  
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai  
201203, P. R. China  
Tel: 86-21-50801919 (ex.45)  
Fax: 86-21-50801922  
Email: hanzg@chgc.sh.cn  
This clone is available at CHGC in Shanghai.

**FEATURES**

**source**  
1. .376  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"

**ORIGIN**

**Query Match** 86.3%; Score 16.4; DB 1; Length 376;  
**Best Local Similarity** 94.4%; Pred. No. 1.7e+03;  
**Matches** 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

**QY** 2 CCAGTGCATCCAAATTGA 19  
|||||  
**Db** 232 CCAGTGCATCCAGATTGA 215

**RESULT 36**  
**LOCUS** CL698283/c

**DEFINITION** CL698283 383 bp DNA linear GSS 23-NOV-2004  
SP\_Ba0027L08.r SP\_Ba Sorghum propinquum genomic clone  
SP\_Ba0027L08 3', genomic survey sequence.

**ACCESSION** CL698283  
**VERSION** CL698283.2 GI:55984464  
**KEYWORDS** GSS.  
**SOURCE** Sorghum propinquum  
**ORGANISM** Sorghum propinquum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Sorghum.

**REFERENCE** 1 (bases 1 to 383)  
**AUTHORS** Kim, H., Yu, Y., Wiesotzki, M., Yost, D., Stum, D., Rao, K., Kudrna, D.,  
Muller, C., Soderlund, C., Bowers, J. E., Paterson, A. H. and Wing, R.  
**TITLE** Physical mapping of the sorghum genome  
**JOURNAL** Unpublished (2004)  
**COMMENT** On Nov 23, 2004 this sequence version replaced gi:50264558.

Contact: Rod A. Wing  
Arizona Genomics Institute  
University of Arizona  
Forbes Building Room 303, Tucson, AZ 85721-0036, USA  
Tel: 520 626 9595  
Fax: 520 621 1259  
Email: rwing@genome.arizona.edu  
PCR Primers  
FORWARD: TAA TAC GAC TCA CTA TAG GG  
BACKWARD: CAC TCA TTA GGC ACC CCA  
Place: 0027 row: L column: 08  
Seq primer: CAC TCA TTA GGC ACC CCA  
Class: BAC ends.

**FEATURES**

**source**  
1. .383  
/organism="Sorghum propinquum"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:132711"  
/clone="SP\_Ba0027L08"  
/clone\_lib="SP\_Ba"  
/notes="Vector: pBeloBAC11; Site\_1: HindIII; Site\_2:  
HindIII; Paterson lab BAC library (HindIII)"

**ORIGIN**

**Query Match** 86.3%; Score 16.4; DB 10; Length 383;  
**Best Local Similarity** 94.4%; Pred. No. 1.7e+03;  
**Matches** 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

**QY** 2 CCAGTGCATCCAAATTGA 19  
|||||  
**Db** 99 CCAGTGCATCCAAATTGA 82

**RESULT 37**  
**LOCUS** AV649842/c

386 bp mRNA linear EST 15-JAN-2002



Best Local Similarity 94.4%; Pred. No. 1.7e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CCCAGTGCATCCAAATTG 18  
|||||  
Db 201 CCCAGTGCATCCAAATTG 184

RESULT 40  
CL700771

LOCUS CL700771 400 bp DNA linear GSS 23-NOV-2004  
DEFINITION SP\_Ba0065B22.f SP\_Ba Sorghum propinquum genomic clone  
SP\_Ba0065B22 5', genomic survey sequence.

ACCESSION CL700771  
VERSION CL700771.2 GI:55986706  
KEYWORDS GSS.

SOURCE  
ORGANISM

Sorghum propinquum  
Sorghum propinquum  
Sorghum propinquum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Sorghum.

1 (bases 1 to 400)

AUTHORS Kim H., Yu Y., Wisoraki M., Yost D., Stum D., Rao K., Kudrna D.,  
Muller C., Soderlund C., Bowers J.E., Paterson A.H. and Wing R.

TITLE Physical mapping of the sorghum genome  
JOURNAL Unpublished (2004)

COMMENT On Nov 23, 2004 this sequence version replaced gi:50267046.

Contact: Rod A. Wing  
Arizona Genomics Institute  
University of Arizona  
Forbes Building Room 303, Tucson, AZ 85721-0036, USA  
Tel: 520 626 9595  
Fax: 520 621 1259

Email: rwing@genome.arizona.edu

PCR Primers

FORWARD: TAA TAC GAC TCA CTA TAG GG

BACKWARD: CAC TCA TTA GGC ACC CCA

Plate: 0065 row: B column: 22

Seq primer: TAA TAC GAC TCA CTA TAG GG

Class: BAC ends.

FEATURES

source 1..400  
Location/Qualifiers  
/organism="Sorghum propinquum"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:132711"  
/clone="SP\_Ba0065B22"  
/clone\_lib="SP\_Ba"  
/notes="Vector: pBelobAC11; Site\_1: HindIII; Site\_2:  
HindIII; Paterson lab BAC library (HindIII)"

ORIGIN

Query Match 86.3%; Score 16.4; DB 10; Length 400;  
Best Local Similarity 94.4%; Pred. No. 1.7e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 CCAGTGCATCCAAATTGA 19  
|||||  
Db 199 CCAGTGCATCCACATTGA 216

Search completed: November 20, 2005, 21:55:35  
Job time : 2844.56 secs

This Page Blank (uspto)



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2005, 17:14:50 ; Search time 86.5555 Seconds  
(without alignments)  
390.196 Million cell updates/sec

Title: US-10-627-757-27

Perfect score: 19  
Sequence: 1 ccagtgatcccaattga 19

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA:\*  
1: /cgn2\_6/prodata/1/ina/1 COMB.seq:\*  
2: /cgn2\_6/prodata/1/ina/5 COMB.seq:\*  
3: /cgn2\_6/prodata/1/ina/6A COMB.seq:\*  
4: /cgn2\_6/prodata/1/ina/6B COMB.seq:\*  
5: /cgn2\_6/prodata/1/ina/H COMB.seq:\*  
6: /cgn2\_6/prodata/1/ina/PCUS COMB.seq:\*  
7: /cgn2\_6/prodata/1/ina/PP COMB.seq:\*  
8: /cgn2\_6/prodata/1/ina/RE COMB.seq:\*  
9: /cgn2\_6/prodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	16.4	86.3	601	3	US-09-949-016-27213 Sequence 27213, A
2	16.4	86.3	601	3	US-09-949-016-27214 Sequence 27214, A
3	16.4	86.3	601	3	US-09-949-016-27215 Sequence 27215, A
4	16.4	86.3	601	3	US-09-949-016-78083 Sequence 78083, A
5	16.4	86.3	601	3	US-09-949-016-78084 Sequence 78084, A
6	16.4	86.3	601	3	US-09-949-016-78085 Sequence 78085, A
7	16.4	86.3	1404	3	US-09-543-681A-3255 Sequence 3255, A
8	16.4	86.3	2585	3	US-09-949-016-3121 Sequence 3121, A
9	16.4	86.3	2878	3	US-10-108-767-19 Sequence 19, Appl
10	16.4	86.3	2884	3	US-09-949-016-2279 Sequence 2729, A
11	16.4	86.3	2900	3	US-09-949-016-474 Sequence 474, A
12	16.4	86.3	25227	3	US-09-949-016-11847 Sequence 11847, A
13	16.4	86.3	25227	3	US-09-949-016-14863 Sequence 14863, A
14	16.4	86.3	36759	3	US-09-949-016-12216 Sequence 12216, A
15	16.4	86.3	36760	3	US-09-949-016-14021 Sequence 14021, A
16	15.8	83.2	3625	3	US-09-023-655-1180 Sequence 1180, A
17	15.4	81.1	387	3	US-09-513-999C-1611 Sequence 1611, A
18	15.4	81.1	444	3	US-09-621-976-13135 Sequence 13135, A
19	15.4	81.1	473	3	US-09-621-976-13130 Sequence 13130, A
20	15.4	81.1	601	3	US-09-949-016-167038 Sequence 167038, A
21	15.4	81.1	601	3	US-09-949-016-203250 Sequence 203250, A
22	15.4	81.1	699	3	US-09-270-767-192 Sequence 192, A
23	15.4	81.1	699	3	US-09-270-767-15474 Sequence 15474, A
24	15.4	81.1	1817	3	US-09-949-016-5735 Sequence 5735, A

25	15.4	81.1	1850	3	US-09-566-921-129 Sequence 129, A
26	15.4	81.1	3438	3	US-10-104-047-559 Sequence 559, A
27	15.4	81.1	27589	3	US-09-949-016-17477 Sequence 17477, A
28	15.4	81.1	32172	3	US-09-949-016-16440 Sequence 16440, A
29	15	78.9	211049	3	US-09-949-016-15770 Sequence 15770, A
30	14.8	77.9	291	3	US-09-621-976-2031 Sequence 2031, A
31	14.8	77.9	601	3	US-09-949-016-31230 Sequence 31230, A
32	14.8	77.9	601	3	US-09-949-016-81354 Sequence 81354, A
33	14.8	77.9	864	3	US-09-107-532A-679 Sequence 679, A
34	14.8	77.9	921	4	US-09-605-703B-2391 Sequence 2391, A
35	14.8	77.9	1083	3	US-09-107-532A-3012 Sequence 3012, A
36	14.8	77.9	1497	3	US-09-107-532A-1428 Sequence 1428, A
37	14.8	77.9	1857	3	US-09-134-000C-3133 Sequence 3133, A
38	14.8	77.9	2165	2	US-08-669-721-2 Sequence 2, Appli
39	14.8	77.9	2165	3	US-09-189-344-2 Sequence 2, Appli
40	14.8	77.9	2165	4	US-09-839-743-2 Sequence 2, Appli
41	14.8	77.9	2495	4	US-09-605-703B-357 Sequence 357, A
42	14.8	77.9	2576	3	US-09-949-016-105 Sequence 105, A
43	14.8	77.9	2619	4	US-09-605-703B-355 Sequence 355, A
44	14.8	77.9	8779	2	US-08-750-703-4 Sequence 4, Appli
45	14.8	77.9	15412	3	US-09-949-016-15869 Sequence 15869, A

ALIGNMENTS

RESULT 1  
US-09-949-016-27213  
; Sequence 27213, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 27213  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-27213

Query Match 86.3%; Score 16.4; DB 3; Length 601;  
Best Local Similarity 94.4%; Pred. No. 57;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATTG 18  
|||||  
DB 11 CCCAGTGCATCCAAATG 28

RESULT 2  
US-09-949-016-27214  
; Sequence 27214, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20

```
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27214
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-27214
```

```
Query Match      86.3%; Score 16.4; DB 3; Length 601;
Best Local Similarity 94.4%; Pred. No. 57;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 CCCAGTGCATCCAAATTG 18
    |||||
Db 239 CCCAGTGCATCCAAATG 256
```

## RESULT 3

```
US-09-949-016-27215
; Sequence 27215, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27215
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-27215
```

```
Query Match      86.3%; Score 16.4; DB 3; Length 601;
Best Local Similarity 94.4%; Pred. No. 57;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 CCCAGTGCATCCAAATTG 18
    |||||
Db 240 CCCAGTGCATCCAAATG 257
```

## RESULT 4

```
US-09-949-016-78083
; Sequence 78083, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 78083
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-78083
```

```
Query Match      86.3%; Score 16.4; DB 3; Length 601;
Best Local Similarity 94.4%; Pred. No. 57;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 CCCAGTGCATCCAAATTG 18
    |||||
Db 11 CCCAGTGCATCCAAATG 28
```

## RESULT 5

```
US-09-949-016-78084
; Sequence 78084, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 78084
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-78084
```

```
Query Match      86.3%; Score 16.4; DB 3; Length 601;
Best Local Similarity 94.4%; Pred. No. 57;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 CCCAGTGCATCCAAATTG 18
    |||||
Db 239 CCCAGTGCATCCAAATG 256
```

## RESULT 6

```
US-09-949-016-78085
; Sequence 78085, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 78085
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
```

US-09-949-016-78085

Query Match 86.3%; Score 16.4; DB 3; Length 601;  
Best Local Similarity 94.4%; Pred. No. 57;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATTG 18  
|||||  
Db 240 CCCAGTGCATCCAAATG 257

RESULT 7

US-09-543-681A-3255/c  
; Sequence 3255, Application US/09543681A  
; Patent No. 6605709

; GENERAL INFORMATION:

; APPLICANT: GARY BRETON

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS

; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709,1002-001

; CURRENT APPLICATION NUMBER: US/09/543,681A

; CURRENT FILING DATE: 2000-04-05

; PRIOR APPLICATION NUMBER: US 60/128,706

; PRIOR FILING DATE: 1999-04-09

; NUMBER OF SEQ ID NOS: 8344

; SEQ ID NO 3255

; LENGTH: 1404

; TYPE: DNA

; ORGANISM: Proteus mirabilis

US-09-543-681A-3255

Query Match 86.3%; Score 16.4; DB 3; Length 1404;  
Best Local Similarity 94.4%; Pred. No. 64;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATTG 18  
|||||  
Db 824 CCCAGTGCAGCCAAATTG 807

RESULT 8

US-09-949-016-3121/c

; Sequence 3121, Application US/09949016

; Patent No. 6812339

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001307

; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498

; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 3121

; LENGTH: 2585

; TYPE: DNA

; ORGANISM: Human

US-09-949-016-3121

Query Match 86.3%; Score 16.4; DB 3; Length 2585;  
Best Local Similarity 94.4%; Pred. No. 69;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCAGTGCATCCAAATTGA 19  
|||||  
Db 1538 CCAGTGCATCCAAATTGA 1521

RESULT 9

US-10-108-767-19/c

; Sequence 19, Application US/10108767

; Patent No. 6919184

; GENERAL INFORMATION:

; APPLICANT: Issakani, Sarkiz D.

; APPLICANT: Huang, Jianing

; APPLICANT: Sheung, Julie

; APPLICANT: Pray, Todd R.

; TITLE OF INVENTION: ASSAYS FOR IDENTIFYING UBIQUITIN AGENTS AND FOR IDENTIFYING AGENTS

; TITLE OF INVENTION: MODIFY THE ACTIVITY OF UBIQUITIN AGENTS

; FILE REFERENCE: A-68613-5/RMS/DCF

; CURRENT APPLICATION NUMBER: US/10/108,767

; CURRENT FILING DATE: 2002-09-26

; PRIOR APPLICATION NUMBER: US 09/542,497

; PRIOR FILING DATE: 2000-04-03

; PRIOR APPLICATION NUMBER: US 09/826,312

; PRIOR FILING DATE: 2001-04-03

; PRIOR APPLICATION NUMBER: US 10/091,139

; PRIOR FILING DATE: 2002-03-04

; NUMBER OF SEQ ID NOS: 27

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 19

; LENGTH: 2878

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-108-767-19

Query Match 86.3%; Score 16.4; DB 3; Length 2878;  
Best Local Similarity 94.4%; Pred. No. 70;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATTG 18  
|||||  
Db 1714 CCCAGTGCATCCAAATG 1697

RESULT 10

US-09-949-016-2279/c

; Sequence 2279, Application US/09949016

; Patent No. 6812339

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001307

; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498

; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2279

; LENGTH: 2884

; TYPE: DNA

; ORGANISM: Human

US-09-949-016-2279

Query Match 86.3%; Score 16.4; DB 3; Length 2884;  
Best Local Similarity 94.4%; Pred. No. 70;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATTG 18  
|||||  
Db 1719 CCCAGTGCATCCAAATG 1702

RESULT 11

```
US-09-949-016-474/c
; Sequence 474, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 474
; LENGTH: 2900
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(2900)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-474

Query Match      86.3%; Score 16.4; DB 3; Length 2900;
Best Local Similarity 94.4%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CCAGTGCATCCAAATTG 18
Db      1725 CCAGTGCATCCAAATG 1708

RESULT 12
US-09-949-016-11847/c
; Sequence 11847, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11847
; LENGTH: 25227
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-11847

Query Match      86.3%; Score 16.4; DB 3; Length 25227;
Best Local Similarity 94.4%; Pred. No. 91;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 CCAGTGCATCCAAATTG 19
Db      20050 CCAGTGCATCCAGATTG 20033

RESULT 13
US-09-949-016-14863/c
; Sequence 14863, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14863
; LENGTH: 25227
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14863

Query Match      86.3%; Score 16.4; DB 3; Length 25227;
Best Local Similarity 94.4%; Pred. No. 91;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 CCAGTGCATCCAAATTG 19
Db      20050 CCAGTGCATCCAGATTG 20033

RESULT 14
US-09-949-016-12216/c
; Sequence 12216, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12216
; LENGTH: 36759
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12216

Query Match      86.3%; Score 16.4; DB 3; Length 36759;
Best Local Similarity 94.4%; Pred. No. 95;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CCAGTGCATCCAAATTG 18
Db      33595 CCAGTGCATCCAAATG 33578

RESULT 15
US-09-949-016-14021/c
; Sequence 14021, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
```

```
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14021
; LENGTH: 36760
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14021

Query Match      86.3%; Score 16.4; DB 3; Length 36760;
Best Local Similarity 94.4%; Pred. No. 95;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CCCAGTGCATCCAAATTG 18
Db      33595 CCCAGTGCATCCAAATG 33578

RESULT 16
US-09-023-655-1180
; Sequence 1180, Application US/09023655
; Patent No. 6607879
; GENERAL INFORMATION:
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Susan G. Stuart
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
; TITLE OF INVENTION: EXPRESSION
; NUMBER OF SEQUENCES: 1508
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,655
; FILING DATE: HEREWTH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0001 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1180:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3625 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
```

```
; CLONE: g189177
US-09-023-655-1180

Query Match      83.2%; Score 15.8; DB 3; Length 3625;
Best Local Similarity 89.5%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CCCAGTGCATCCAAATTGA 19
Db      1130 CCCAATGCATCCCAACTTGA 1148

RESULT 17
US-09-513-999C-1611
; Sequence 1611, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duciart, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59 US2 REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 1611
; LENGTH: 387
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 59..385
US-09-513-999C-1611

Query Match      81.1%; Score 15.4; DB 3; Length 387;
Best Local Similarity 94.1%; Pred. No. 1.8e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CCCAGTGCATCCAAATT 17
Db      367 CCCAATGCATCCCAATT 383

RESULT 18
US-09-621-976-13135
; Sequence 13135, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 13135
; LENGTH: 444
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-13135

Query Match      81.1%; Score 15.4; DB 3; Length 444;
Best Local Similarity 94.1%; Pred. No. 1.8e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 CAGTGCATCCCAATTGA 19
Db      1130 CCCAATGCATCCCAACTTGA 1148
```

```
Db      179 CAGTGGATCCAAATTGA 195
;
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 203250
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-203250

Query Match      81.1%; Score 15.4; DB 3; Length 601;
Best Local Similarity 94.1%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CCAGTGCATCCAAATT 17
Db      471 CCCAATGCATCCAAATT 455

RESULT 22
US-09-270-767-192
; Sequence 192, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 192
; LENGTH: 699
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-192

Query Match      81.1%; Score 15.4; DB 3; Length 699;
Best Local Similarity 94.1%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 CAGTGCATCCAAATTGA 19
Db      584 CAGCGCATCCAAATTGA 600

RESULT 23
US-09-270-767-15474
; Sequence 15474, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15474
; LENGTH: 699
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-15474

Query Match      81.1%; Score 15.4; DB 3; Length 699;
;
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 167038
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-167038

Query Match      81.1%; Score 15.4; DB 3; Length 601;
Best Local Similarity 94.1%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CCAGTGCATCCAAATT 17
Db      409 CCCAGTGCATCCCAATT 393

RESULT 21
US-09-949-016-203250/c
; Sequence 203250, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
```

```
Best Local Similarity 94.1%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 CAGTGCATCCAAATTGA 19
   ||| ||||| ||||| |||||
Db 584 CAGCGCATCCAAATTGA 600

RESULT 24
US-09-949-016-5735
; Sequence 5735, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5735
; LENGTH: 1817
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-5735

Query Match 81.1%; Score 15.4; DB 3; Length 1817;
Best Local Similarity 94.1%; Pred. No. 2.2e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATT 17
   ||| ||||| ||||| |||||
Db 364 CCCAATGCATCCAAATT 380

RESULT 25
US-09-566-921-129
; Sequence 129, Application US/09566921
; Patent No. 682888
; GENERAL INFORMATION:
; APPLICANT: Loring, Jeanne F.
; APPLICANT: Tingley, Debra W.
; APPLICANT: Edwards, Carla M.
; TITLE OF INVENTION: GENES EXPRESSED IN ALZHEIMER'S DISEASE
; FILE REFERENCE: PA-0024 US
; CURRENT APPLICATION NUMBER: US/09/566,921
; CURRENT FILING DATE: 2000-05-05
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: PERL Program
; SEQ ID NO 129
; LENGTH: 1850
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 682888 1095628.1
US-09-566-921-129

Query Match 81.1%; Score 15.4; DB 3; Length 1850;
Best Local Similarity 94.1%; Pred. No. 2.2e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATT 17
   ||| ||||| ||||| |||||
Db 396 CCCAATGCATCCAAATT 412
```

```
RESULT 26
US-10-104-047-559/c
; Sequence 559, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. 6943241el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 559
; LENGTH: 3438
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-559

Query Match 81.1%; Score 15.4; DB 3; Length 3438;
Best Local Similarity 94.1%; Pred. No. 2.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 CAGTGCATCCAAATTGA 19
   ||| ||||| ||||| |||||
Db 2296 CAGTGGTCCAAATTGA 2280

RESULT 27
US-09-949-016-17477
; Sequence 17477, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17477
; LENGTH: 27589
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-17477

Query Match 81.1%; Score 15.4; DB 3; Length 27589;
Best Local Similarity 94.1%; Pred. No. 3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATT 17
   ||| ||||| ||||| |||||
Db 20166 CCCAATGCATCCAAATT 20182

RESULT 28
US-09-949-016-16440/c
; Sequence 16440, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
```

```
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16440
; LENGTH: 32172
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(32172)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16440

Query Match      81.1%; Score 15.4; DB 3; Length 32172;
Best Local Similarity 94.1%; Pred. No. 3.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CCCAGTGCATCCAAATT 17
Db      20335 CCCAGTGCATCCACATT 20319

RESULT 29
US-09-949-016-15770/c
; Sequence 15770, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15770
; LENGTH: 211049
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15770

Query Match      78.9%; Score 15; DB 3; Length 211049;
Best Local Similarity 100.0%; Pred. No. 6.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      4 AGTGCATCCAAATTG 18
Db      72974 AGTGCATCCAAATTG 72960

RESULT 30
US-09-621-976-2031
; Sequence 2031, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2

; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 2031
; LENGTH: 291
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 130..288
; NAME/KEY: sig_peptide
; LOCATION: 130..234
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 3.5
; OTHER INFORMATION: seq GLVEVAVGVVRS/DQ
US-09-621-976-2031

Query Match      77.9%; Score 14.8; DB 3; Length 291;
Best Local Similarity 88.9%; Pred. No. 3.6e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CCCAGTGCATCCAAATTG 18
Db      22 CCCAGTGCACCACTTG 39

RESULT 31
US-09-949-016-31230
; Sequence 31230, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31230
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-31230

Query Match      77.9%; Score 14.8; DB 3; Length 601;
Best Local Similarity 88.9%; Pred. No. 3.9e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CCCAGTGCATCCAAATTG 18
Db      512 CCCAGTGCCTCCAAATTG 529

RESULT 32
US-09-949-016-81354
; Sequence 81354, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
```



```

; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 81354
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-81354

Query Match      77.9%; Score 14.8; DB 3; Length 601;
Best Local Similarity 88.9%; Pred. No. 3.9e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CCCAGTGCATCCAAATTG 18
Db      512 CCCAGTGCATCCAAATTG 529

RESULT 33
US-09-107-532A-679/c
; Sequence 679, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 679:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 864 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...864
; SEQUENCE DESCRIPTION: SEQ ID NO: 679:

```

```

US-09-107-532A-679

Query Match      77.9%; Score 14.8; DB 3; Length 864;
Best Local Similarity 88.9%; Pred. No. 4.1e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CCCAGTGCATCCAAATTG 18
Db      517 CACAGTGCATCCAAATTG 500

RESULT 34
US-09-605-703B-2391
; Sequence 2391, Application US/09605703B
; Patent No. 6962989
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
; APPLICANT: Kroeger, Burkhard
; APPLICANT: Schroeder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberhauer, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING NOVEL
; TITLE OF INVENTION: PROTEINS
; FILE REFERENCE: BGI-129CP
; CURRENT APPLICATION NUMBER: US/09/605,703B
; CURRENT FILING DATE: 2000-06-27
; PRIOR APPLICATION NUMBER: 60/142,764
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: 60/152,318
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 2934
; SEQ ID NO 2391
; LENGTH: 921
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (101)..(898)
; OTHER INFORMATION: RXA01741
US-09-605-703B-2391

Query Match      77.9%; Score 14.8; DB 4; Length 921;
Best Local Similarity 88.9%; Pred. No. 4.1e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CCCAGTGCATCCAAATTG 18
Db      856 CCTGTGCATCCAAATG 873

RESULT 35
US-09-107-532A-3012/c
; Sequence 3012, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 679:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 864 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...864
; SEQUENCE DESCRIPTION: SEQ ID NO: 679:

```

```
;
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3012:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1083 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...1083
; SEQUENCE DESCRIPTION: SEQ ID NO: 3012:
US-09-107-532A-3012

Query Match 77.9%; Score 14.8; DB 3; Length 1083;
Best Local Similarity 88.9%; Pred. No. 4.2e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CCAGTGCATCCAAATTGA 19
   |||||
Db 898 CCAGTGCCTTCAAAATTGA 881

RESULT 36
US-09-107-532A-1428/c
; Sequence 1428, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A. Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; CORRESPONDENCE ADDRESS:
; ADDRESS: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
```

```
;
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 1428:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1497 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...1497
; SEQUENCE DESCRIPTION: SEQ ID NO: 1428:
US-09-107-532A-1428

Query Match 77.9%; Score 14.8; DB 3; Length 1497;
Best Local Similarity 88.9%; Pred. No. 4.4e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CCAGTGCATCCAAATTGA 19
   |||||
Db 262 CCATGCGATCCAAATTGA 245

RESULT 37
US-09-134-000C-3133/c
; Sequence 3133, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3133
; LENGTH: 1857
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-134-000C-3133

Query Match 77.9%; Score 14.8; DB 3; Length 1857;
Best Local Similarity 88.9%; Pred. No. 4.5e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCAGTGCATCCAAATTG 18
   |||||
Db 155 CCAGTGCCTTCAAAATTG 138

RESULT 38
US-08-669-721-2
; Sequence 2, Application US/08669721
; Patent No. 5834236
; GENERAL INFORMATION:
; APPLICANT: Lamb et al., Christopher J.
; TITLE OF INVENTION: NOVEL TRANSCRIPTION ENHANCER ELEMENT AND
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
```

```
;
;
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 27-JUN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ellison, Eldora L.
; REGISTRATION NUMBER: 39,967
; REFERENCE/DOCKET NUMBER: 07251/014001
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2165 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 61..1698
; US-08-669-721-2

Query Match 77.9%; Score 14.8; DB 2; Length 2165;
Best Local Similarity 88.9%; Pred. No. 4.6e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CCAGTGCATCCAAATTGA 19
Db 1684 CCAGTGCAGCAAAATTGA 1701

RESULT 39
US-09-189-344-2
; Sequence 2, Application US/09189344
; Patent No. 6191258
; GENERAL INFORMATION:
; APPLICANT: Lamb et al., Christopher J.
; TITLE OF INVENTION: NOVEL TRANSCRIPTION ENHANCER ELEMENT AND
; TITLE OF INVENTION: TRANSCRIPTION FACTOR AND METHODS OF USE THEREFOR
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION NUMBER: US 08/669,721
; FILING DATE: 27-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ellison, Eldora L.
; REGISTRATION NUMBER: 39,967
; REFERENCE/DOCKET NUMBER: 07251/014001
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 2:
```

```
;
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2165 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 61..1698
; US-09-189-344-2

Query Match 77.9%; Score 14.8; DB 3; Length 2165;
Best Local Similarity 88.9%; Pred. No. 4.6e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CCAGTGCATCCAAATTGA 19
Db 1684 CCAGTGCAGCAAAATTGA 1701

RESULT 40
US-09-839-743-2
; Sequence 2, Application US/09839743
; Patent No. 6472211
; GENERAL INFORMATION:
; APPLICANT: The Salk Institute for Biological Sciences
; APPLICANT: Lamb, Christopher
; APPLICANT: Doerner, Peter
; APPLICANT: Laible, Goetz
; TITLE OF INVENTION: No. 6472211el Transcription Enhancer Element and
; TITLE OF INVENTION: Transcription Factor and Methods of Use Thereof
; FILE REFERENCE: SALKINS.008DV3
; CURRENT APPLICATION NUMBER: US/09/839,743
; PRIOR FILING DATE: 2001-04-19
; PRIOR APPLICATION NUMBER: US 09/401,336
; PRIOR FILING DATE: 1999-09-21
; PRIOR APPLICATION NUMBER: US 09/189,344
; PRIOR FILING DATE: 1998-11-10
; PRIOR APPLICATION NUMBER: US 08/669,721
; PRIOR FILING DATE: 1996-06-27
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 2165
; TYPE: DNA
; ORGANISM: Nicotiana tabacum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (61)...(1698)
; US-09-839-743-2

Query Match 77.9%; Score 14.8; DB 3; Length 2165;
Best Local Similarity 88.9%; Pred. No. 4.6e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CCAGTGCATCCAAATTGA 19
Db 1684 CCAGTGCAGCAAAATTGA 1701
```

Search completed: November 20, 2005, 22:02:08  
Job time : 87.5555 secs

**This Page Blank (uspto)**

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2005, 18:34:25 ; Search time 572.346 Seconds  
(without alignments)  
274.516 Million cell updates/sec

Title: US-10-627-757-27

Perfect score: 19

Sequence: 1 ccacgtgcatacaattga 19

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 413469005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA Main:\*

- 1: /cgn2\_6/prodata/1/pubpna/us07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/prodata/1/pubpna/us08\_PUBCOMB.seq:\*
- 3: /cgn2\_6/prodata/1/pubpna/us09\_PUBCOMB.seq:\*
- 4: /cgn2\_6/prodata/1/pubpna/us09\_PUBCOMB.seq:\*
- 5: /cgn2\_6/prodata/1/pubpna/us10\_PUBCOMB.seq:\*
- 6: /cgn2\_6/prodata/1/pubpna/us10\_PUBCOMB.seq:\*
- 7: /cgn2\_6/prodata/1/pubpna/us10\_PUBCOMB.seq:\*
- 8: /cgn2\_6/prodata/1/pubpna/us10\_PUBCOMB.seq:\*
- 9: /cgn2\_6/prodata/1/pubpna/us10\_PUBCOMB.seq:\*
- 10: /cgn2\_6/prodata/1/pubpna/us11\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19	100.0	19	7	US-10-627-757-27
2	19	100.0	623	4	US-09-925-065A-684170
3	19	100.0	1116	7	US-10-627-757-8
4	19	100.0	46951	6	US-10-091-281-2
5	17.4	91.6	186	7	US-10-424-599-72624
6	17	89.5	595	4	US-09-925-065A-788974
7	16.4	86.3	465	3	US-09-880-107-17
8	16.4	86.3	485	7	US-10-437-963-50524
9	16.4	86.3	561	4	US-09-925-065A-219659
10	16.4	86.3	569	4	US-09-925-065A-157780
11	16.4	86.3	595	4	US-09-925-065A-489597
12	16.4	86.3	711	10	US-11-097-143-35773
13	16.4	86.3	1721	6	US-10-006-285-438
14	16.4	86.3	2711	10	US-11-097-143-39772
15	16.4	86.3	2878	5	US-10-108-767-19
16	16.4	86.3	2878	5	US-10-152-156-19
17	16.4	86.3	2900	5	US-10-205-823-428
18	16.4	86.3	2900	10	US-11-051-454-428
19	16.4	86.3	2919	9	US-10-887-553A-552
20	16.4	86.3	150437	9	US-10-981-277-50
21	16	84.2	445	9	US-10-915-740A-760
22	16	84.2	564	9	US-10-972-079-87794
23	16	84.2	589	4	US-09-925-065A-248192

c 24	16	84.2	589	4	US-09-925-065A-248193	Sequence 248193,
c 25	16	84.2	592	9	US-10-972-079-87793	Sequence 87793, A
c 26	16	84.2	600	9	US-10-972-079-87791	Sequence 87791, A
c 27	16	84.2	600	9	US-10-972-079-87792	Sequence 87792, A
c 28	15.8	83.2	512	3	US-09-864-761-7436	Sequence 7436, Ap
c 29	15.8	83.2	529	4	US-09-925-065A-190164	Sequence 190164,
c 30	15.8	83.2	582	9	US-10-972-079-73299	Sequence 73299, A
c 31	15.8	83.2	586	4	US-09-925-065A-918407	Sequence 918407,
c 32	15.8	83.2	599	9	US-10-972-079-25977	Sequence 25977, A
c 33	15.8	83.2	600	9	US-10-972-079-25978	Sequence 25978, A
c 34	15.8	83.2	600	9	US-10-972-079-25979	Sequence 25979, A
c 35	15.8	83.2	600	9	US-10-972-079-73298	Sequence 73298, A
c 36	15.8	83.2	2538	9	US-10-960-409-5	Sequence 5, Appl
c 37	15.8	83.2	2811	9	US-10-960-409-3	Sequence 3, Appl
c 38	15.8	83.2	2907	7	US-10-394-948-30	Sequence 30, Appl
c 39	15.8	83.2	2907	7	US-10-052-482-120	Sequence 120, App
c 40	15.8	83.2	3497	6	US-10-247-671-22	Sequence 22, Appl
c 41	15.8	83.2	3625	6	US-10-101-510-122	Sequence 122, App
c 42	15.8	83.2	3625	6	US-10-392-113-22	Sequence 22, Appl
c 43	15.8	83.2	3625	6	US-10-189-256-4	Sequence 4, Appl
c 44	15.8	83.2	3625	7	US-10-394-948-29	Sequence 29, Appl
c 45	15.8	83.2	3625	7	US-10-052-482-119	Sequence 119, App

#### ALIGNMENTS

##### RESULT 1

US-10-627-757-27

; Sequence 27, Application US/10627757

; Publication No. US20040091914A1

; GENERAL INFORMATION:

; APPLICANT: KOUCHI YASUHIRO

; APPLICANT: MASASATO AKINORI

; APPLICANT: TAKAHITO TAKAYUKI

; TITLE OF INVENTION: GENE ASSAY METHOD FOR PREDICTING GLAUCOMA ONSET RISK

; FILE REFERENCE: Q76319

; CURRENT APPLICATION NUMBER: US/10/627,757

; CURRENT FILING DATE: 2003-07-28

; PRIOR APPLICATION NUMBER: JP P2002-226612

; PRIOR FILING DATE: 2002-08-02

; NUMBER OF SEQ ID NOS: 40

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 27

; LENGTH: 19

; TYPE: DNA

; ORGANISM: Artificial

; FEATURE:

; OTHER INFORMATION: Designed DNA based on OPTN gene

US-10-627-757-27

Query Match 100.0%; Score 19; DB 7; Length 19;

Best Local Similarity 100.0%; Pred. No. 9.6;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCCAGTGCATCCAAATTGA 19

Db 1 CCCAGTGCATCCAAATTGA 19

##### RESULT 2

US-09-925-065A-684170/c

; Sequence 684170, Application US/09925065A

; Publication No. US20050228172A9

; GENERAL INFORMATION:

; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single

; FILE REFERENCE: 108827.135

; CURRENT APPLICATION NUMBER: US/09/925,065A

; CURRENT FILING DATE: 2001-08-08

; PRIOR APPLICATION NUMBER: US 60/243,096

; PRIOR FILING DATE: 2000-10-24

```
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 684170
; LENGTH: 623
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-684170

Query Match      100.0%; Score 19; DB 4; Length 623;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CCCAGTGCATCCAAATTGA 19
        |||||
DB      127 CCCAGTGCATCCAAATTGA 109

RESULT 3
US-10-627-757-8
; Sequence 8, Application US/10627757
; Publication No. US20040091914A1
; GENERAL INFORMATION:
; APPLICANT: KOUCHI YASUHIRO
; APPLICANT: MASASGO AKINORI
; APPLICANT: TAKAHATI TAKAYUKI
; TITLE OF INVENTION: GENE ASSAY METHOD FOR PREDICTING GLAUCOMA ONSET RISK
; FILE REFERENCE: Q76319
; CURRENT APPLICATION NUMBER: US/10/627,757
; CURRENT FILING DATE: 2003-07-28
; PRIOR APPLICATION NUMBER: JP P2002-226612
; PRIOR FILING DATE: 2002-08-02
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 1116
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-627-757-8

Query Match      100.0%; Score 19; DB 7; Length 1116;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CCCAGTGCATCCAAATTGA 19
        |||||
DB      251 CCCAGTGCATCCAAATTGA 269

RESULT 4
US-10-091-281-2
; Sequence 2, Application US/10091281
; Publication No. US20030190617A1
; GENERAL INFORMATION:
; APPLICANT: RAYMOND, VINCENT
; APPLICANT: SI, ERWIN
; APPLICANT: MORISSETTE, JEAN
; TITLE OF INVENTION: OPTINEURIN NUCLEIC ACID MOLECULES AND USES THEREOF
; FILE REFERENCE: 13587.338
; CURRENT APPLICATION NUMBER: US/10/091,281
; CURRENT FILING DATE: 2002-03-06
; NUMBER OF SEQ ID NOS: 463
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 46951
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: allele
; LOCATION: 191
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 691
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 709
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 887
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 894
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 987
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 1112
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 1505
; OTHER INFORMATION: insertion of additional c residue
; NAME/KEY: allele
; LOCATION: 1606
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 2405
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 2606
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 3313
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 3555
; OTHER INFORMATION: insertion of additional t residue
; NAME/KEY: allele
; LOCATION: 3625
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 3629
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 3882
; OTHER INFORMATION: insertion of additional t residue
; NAME/KEY: allele
; LOCATION: 3988
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 4452
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: repeat_region
; LOCATION: 598..878
; OTHER INFORMATION: repeat element
; NAME/KEY: repeat_region
; LOCATION: 938..957
; OTHER INFORMATION: Short repeat element
; NAME/KEY: repeat_region
; LOCATION: 1002..1329
; OTHER INFORMATION: ALU repeat element
; NAME/KEY: repeat_region
; LOCATION: 2288..2587
; OTHER INFORMATION: ALU repeat element
; NAME/KEY: misc_feature
; LOCATION: 5054
; OTHER INFORMATION: putative transcription start site
US-10-091-281-2
```

Query Match 100.0%; Score 19; DB 6; Length 46951;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCAGTGCATCCAAATTGA 19  
|||||  
Db 28589 CCAGTGCATCCAAATTGA 28607

RESULT 5  
US-10-424-599-72624  
; Sequence 72624, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 72624  
; LENGTH: 186  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(186)  
; OTHER INFORMATION: unsure at all n locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_36592C.1  
US-10-424-599-72624

Query Match 91.8%; Score 17.4; DB 7; Length 186;  
Best Local Similarity 94.7%; Pred. No. 82;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCAGTGCATCCAAATTGA 19  
|||||  
Db 113 CCAGTGCATCCACATTGA 131

RESULT 6  
US-09-925-065A-788974/c  
; Sequence 788974, Application US/09925065A  
; Publication No. US2005028172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 788974  
; LENGTH: 595  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-788974

Query Match 89.5%; Score 17; DB 4; Length 595;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCAGTGCATCCAAATT 17  
|||||  
Db 379 CCAGTGCATCCAAATT 363

RESULT 7  
US-09-880-107-17  
; Sequence 17, Application US/09880107  
; Patent No. US20020142981A1  
; GENERAL INFORMATION:  
; APPLICANT: Vockley, Darci T.  
; APPLICANT: Scherf, Uwe  
; APPLICANT: Gene Logic, Inc.  
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer  
; FILE REFERENCE: 44921-5028-WO  
; CURRENT APPLICATION NUMBER: US/09/880,107  
; CURRENT FILING DATE: 2001-06-14  
; PRIOR APPLICATION NUMBER: US 60/211,379  
; PRIOR FILING DATE: 2000-06-14  
; PRIOR APPLICATION NUMBER: US 60/237,054  
; PRIOR FILING DATE: 2000-10-02  
; NUMBER OF SEQ ID NOS: 3950  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 17  
; LENGTH: 465  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 AA007395  
US-09-880-107-17

Query Match 86.3%; Score 16.4; DB 3; Length 465;  
Best Local Similarity 94.4%; Pred. No. 3e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCAGTGCATCCAAATTGA 19  
|||||  
Db 398 CCAGTGCATCCAGATTGA 415

RESULT 8  
US-10-437-963-50524  
; Sequence 50524, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 50524  
; LENGTH: 485  
; TYPE: DNA  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_53001C.1  
US-10-437-963-50524

```
Query Match      86.3%; Score 16.4; DB 7; Length 485;
Best Local Similarity 94.4%; Pred. No. 3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCAGTGCATCCAAATTGA 19
Db 317 CCAGTGCATCCAAATTGA 334

RESULT 9
US-09-925-065A-219659/c
; Sequence 219659, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108927.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; SOFTWARE: FastSeq for Windows Version 4.0
; NUMBER OF SEQ ID NOS: 957086
; SEQ ID NO 219659
; LENGTH: 561
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-219659

Query Match      86.3%; Score 16.4; DB 4; Length 561;
Best Local Similarity 94.4%; Pred. No. 3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCAGTGCATCCAAATTG 18
Db 473 CCAGTGCATCCAAATTG 456

RESULT 10
US-09-925-065A-157780/c
; Sequence 157780, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108927.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 157780
; LENGTH: 569
; TYPE: DNA
US-09-925-065A-157780/c

Query Match      86.3%; Score 16.4; DB 7; Length 485;
Best Local Similarity 94.4%; Pred. No. 3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCAGTGCATCCAAATTGA 19
Db 317 CCAGTGCATCCAAATTGA 334

RESULT 9
US-09-925-065A-219659/c
; Sequence 219659, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108927.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 219659
; LENGTH: 561
; TYPE: DNA
US-09-925-065A-219659

Query Match      86.3%; Score 16.4; DB 4; Length 561;
Best Local Similarity 94.4%; Pred. No. 3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCAGTGCATCCAAATTG 18
Db 473 CCAGTGCATCCAAATTG 456

RESULT 10
US-09-925-065A-157780/c
; Sequence 157780, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108927.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 157780
; LENGTH: 569
; TYPE: DNA
US-09-925-065A-157780/c

Query Match      86.3%; Score 16.4; DB 4; Length 595;
Best Local Similarity 94.4%; Pred. No. 3.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCAGTGCATCCAAATTG 18
Db 276 CCAGTGCATCCAAATTG 259

RESULT 12
US-11-097-143-39773/c
; Sequence 39773, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; FILE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
```



; PRIOR APPLICATION NUMBER: 60/175,693  
; PRIOR FILING DATE: 2000-01-12  
; PRIOR APPLICATION NUMBER: 60/184,831  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/191,637  
; PRIOR FILING DATE: 2000-03-23  
; NUMBER OF SEQ ID NOS: 43008  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 39773  
; LENGTH: 711  
; TYPE: DNA  
; ORGANISM: DROSOPHILA  
US-11-097-143-39773

Query Match 86.3%; Score 16.4; DB 10; Length 711;  
Best Local Similarity 94.4%; Pred. No. 3.1e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCAGTGCATCCAAATTGA 19  
Db 334 CCAGTGCATCCAAATCGA 317

## RESULT 13

US-10-006-285-438/c  
; Sequence 438, Application US/10006285  
; Publication No. US20030165854A1  
; GENERAL INFORMATION:  
; APPLICANT: Mary Jane Cunningham  
; APPLICANT: Matthew R. Kaser  
; TITLE OF INVENTION: MARKER GENES RESPONDING TO TREATMENT WITH TOXINS  
; FILE REFERENCE: PA-0039 US  
; CURRENT APPLICATION NUMBER: US/10/006,285  
; CURRENT FILING DATE: 2001-12-05  
; NUMBER OF SEQ ID NOS: 514  
; SOFTWARE: PERL Program  
; SEQ ID NO 438  
; LENGTH: 1721  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20030165854A1 252542.6  
US-10-006-285-438

Query Match 86.3%; Score 16.4; DB 6; Length 1721;  
Best Local Similarity 94.4%; Pred. No. 3.4e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCAGTGCATCCAAATTGA 19  
Db 1112 CCAGTGCATCCAGATTGA 1095

## RESULT 14

US-11-097-143-39772  
; Sequence 39772, Application US/11097143  
; Publication No. US20050208558A1  
; GENERAL INFORMATION:  
; APPLICANT: Venter, J. Craig  
; APPLICANT: et al.  
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID  
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE  
; FILE REFERENCE: CL000728  
; CURRENT APPLICATION NUMBER: US/11/097,143  
; CURRENT FILING DATE: 2005-04-04  
; PRIOR APPLICATION NUMBER: 60/157,832  
; PRIOR FILING DATE: 1999-10-05  
; PRIOR APPLICATION NUMBER: 60/160,191  
; PRIOR FILING DATE: 1999-10-19  
; PRIOR APPLICATION NUMBER: 60/161,932  
; PRIOR FILING DATE: 1999-10-28

; PRIOR APPLICATION NUMBER: 60/164,769  
; PRIOR FILING DATE: 1999-11-12  
; PRIOR APPLICATION NUMBER: 60/173,383  
; PRIOR FILING DATE: 1999-12-28  
; PRIOR APPLICATION NUMBER: 60/175,693  
; PRIOR FILING DATE: 2000-01-12  
; PRIOR APPLICATION NUMBER: 60/184,831  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/191,637  
; PRIOR FILING DATE: 2000-03-23  
; NUMBER OF SEQ ID NOS: 43008  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 39772  
; LENGTH: 2711  
; TYPE: DNA  
; ORGANISM: DROSOPHILA  
US-11-097-143-39772

Query Match 86.3%; Score 16.4; DB 10; Length 2711;  
Best Local Similarity 94.4%; Pred. No. 3.6e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCAGTGCATCCAAATTGA 19  
Db 1378 CCAGTGCATCCAAATCGA 1395

## RESULT 15

US-10-108-767-19/c  
; Sequence 19, Application US/10108767  
; Publication No. US20030104474A1  
; GENERAL INFORMATION:  
; APPLICANT: Issakani, Sarkiz D.  
; APPLICANT: Huang, Jianing  
; APPLICANT: Sheung, Julie  
; APPLICANT: Pray, Todd R.  
; TITLE OF INVENTION: ASSAYS FOR IDENTIFYING UBIQUITIN AGENTS AND FOR IDENTIFYING AGENT  
; FILE REFERENCE: A-68613-5/RMS/DCF  
; CURRENT APPLICATION NUMBER: US/10/108,767  
; CURRENT FILING DATE: 2002-09-26  
; PRIOR APPLICATION NUMBER: US 09/542,497  
; PRIOR FILING DATE: 2000-04-03  
; PRIOR APPLICATION NUMBER: US 09/826,312  
; PRIOR FILING DATE: 2001-04-03  
; PRIOR APPLICATION NUMBER: US 10/091,139  
; PRIOR FILING DATE: 2002-03-04  
; NUMBER OF SEQ ID NOS: 27  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 19  
; LENGTH: 2878  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-108-767-19

Query Match 86.3%; Score 16.4; DB 5; Length 2878;  
Best Local Similarity 94.4%; Pred. No. 3.6e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCAGTGCATCCAAATTG 18  
Db 1714 CCAGTGCATCCAAATG 1697

## RESULT 16

US-10-152-156-19/c  
; Sequence 19, Application US/10152156  
; Publication No. US20030108947A1  
; GENERAL INFORMATION:  
; APPLICANT: Issakani, Sarkiz D.  
; APPLICANT: Huang, Jianing  
; APPLICANT: Sheung, Julie  
; APPLICANT: Pray, Todd R.

```
; TITLE OF INVENTION: ASSAYS FOR IDENTIFYING UBIQUITIN AGENTS AND FOR IDENTIFYING AGENT
; TITLE OF INVENTION: MODIFY THE ACTIVITY OF UBIQUITIN AGENTS
; FILE REFERENCE: A-68613-6/RMS/DCF
; CURRENT APPLICATION NUMBER: US/10/152,156
; CURRENT FILING DATE: 2002-05-20
; PRIOR APPLICATION NUMBER: US 09/542,497
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: US 09/826,312
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 10/091,174
; PRIOR FILING DATE: 2002-03-04
; PRIOR APPLICATION NUMBER: US 10/091,139
; PRIOR FILING DATE: 2002-03-04
; PRIOR APPLICATION NUMBER: US 10/109,460
; PRIOR FILING DATE: 2002-03-26
; PRIOR APPLICATION NUMBER: US 10/108,767
; PRIOR FILING DATE: 2002-03-26
; PRIOR APPLICATION NUMBER: US 60/291,836
; PRIOR FILING DATE: 2001-05-18
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 19
; LENGTH: 2878
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-152-156-19

Query Match      86.3%; Score 16.4; DB 5; Length 2878;
Best Local Similarity 94.4%; Pred. No. 3.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1  CCAGTGCATCCAAATTG 18
Db      1714 CCAGTGCATCCAAATG 1697

RESULT 17
US-10-205-823-428/c
; Sequence 428, Application US/10205823
; Publication No. US20030108963A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Endege, Wilson O.
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Gorbacheva, Bella
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Wonsey, Angela M.
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Anderson, Dustin
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF PROSTATE CANCER
; FILE REFERENCE: MRI-044
; CURRENT APPLICATION NUMBER: US/10/205,823
; CURRENT FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/307,982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314,356
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/325,020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341,746
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362,158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 428
; LENGTH: 2900
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2439, 2441, 2468, 2493
; OTHER INFORMATION: n = A,T,C or G
US-11-051-454-428

Query Match      86.3%; Score 16.4; DB 10; Length 2900;
Best Local Similarity 94.4%; Pred. No. 3.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1  CCAGTGCATCCAAATTG 18
Db      1725 CCAGTGCATCCAAATG 1708

RESULT 18
US-11-051-454-428/c
; Sequence 428, Application US/11051454
; Publication No. US20050191673A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Endege, Wilson O.
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Gorbacheva, Bella
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Wonsey, Angela M.
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Anderson, Dustin
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF PROSTATE CANCER
; FILE REFERENCE: MRI-044
; CURRENT APPLICATION NUMBER: US/11/051,454
; CURRENT FILING DATE: 2005-02-04
; PRIOR APPLICATION NUMBER: US/10/205,823
; PRIOR FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/307,982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314,356
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/325,020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341,746
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362,158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 428
; LENGTH: 2900
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2439, 2441, 2468, 2493
; OTHER INFORMATION: n = A,T,C or G
US-11-051-454-428

Query Match      86.3%; Score 16.4; DB 10; Length 2900;
Best Local Similarity 94.4%; Pred. No. 3.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1  CCAGTGCATCCAAATTG 18
Db      1725 CCAGTGCATCCAAATG 1708

RESULT 19
US-10-887-553A-552/c
```

; Sequence 552, Application US/10987553A  
; Publication No. US20050085436A1  
; GENERAL INFORMATION:  
; APPLICANT: Garza, Dan  
; APPLICANT: Li, Hao  
; TITLE OF INVENTION: Method to treat conditions associated  
; with insulin signalling dysregulation  
; FILE REFERENCE: 4-33262  
; CURRENT APPLICATION NUMBER: US/10/887,553A  
; CURRENT FILING DATE: 2004-07-08  
; PRIOR APPLICATION NUMBER: 60/485,883  
; PRIOR FILING DATE: 2003-08-07  
; NUMBER OF SEQ ID NOS: 1208  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 552  
; LENGTH: 2919  
; TYPE: DNA  
; ORGANISM: human  
US-10-887-553A-552

Query Match 86.3%; Score 16.4; DB 9; Length 2919;  
Best Local Similarity 94.4%; Pred. No. 3.6e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCCAATTG 18  
|||||  
DB 1745 CCCAGTGCATCCCAATTG 1728

RESULT 20  
US-10-981-277-50  
; Sequence 50, Application US/10981277  
; Publication No. US20050181389A1  
; GENERAL INFORMATION:  
; APPLICANT: Davis, Lisa  
; APPLICANT: Harris, Cole  
; TITLE OF INVENTION: Compositions and Methods for Glioma Classification  
; FILE REFERENCE: 03-968-US  
; CURRENT APPLICATION NUMBER: US/10/981,277  
; CURRENT FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: US 60/516,817  
; PRIOR FILING DATE: 2003-11-03  
; NUMBER OF SEQ ID NOS: 57  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 50  
; LENGTH: 150437  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-981-277-50

Query Match 86.3%; Score 16.4; DB 9; Length 150437;  
Best Local Similarity 94.4%; Pred. No. 5.3e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCAGTGCATCCCAATTGA 19  
|||||  
DB 93130 CCAGTGCATCCAGATTGA 93147

RESULT 21  
US-10-915-740A-760/c  
; Sequence 760, Application US/10915740A  
; Publication No. US20050191316A1  
; GENERAL INFORMATION:  
; APPLICANT: Frazer, Claire M.  
; APPLICANT: Hickey, Erin  
; APPLICANT: Peterson, Jeremy  
; APPLICANT: Tettelin, Hervé  
; APPLICANT: Venter, J. Craig  
; APPLICANT: Massignani, Vega  
; APPLICANT: Galeotti, Cesira  
; APPLICANT: Mora, Manroba  
; APPLICANT: Ratti, Giulio

; APPLICANT: Scarselli, Maria  
; APPLICANT: Scarlato, Vincenzo  
; APPLICANT: Rappuoli, Rino  
; APPLICANT: Pizza, Mariagrazia  
; APPLICANT: Grandi, Guido  
; TITLE OF INVENTION: Neisseria Genomic Sequences And Methods Of Their Use  
; FILE REFERENCE: 002441.00090  
; CURRENT APPLICATION NUMBER: US/10/915,740A  
; CURRENT FILING DATE: 2004-08-11  
; PRIOR APPLICATION NUMBER: 09/806,866  
; PRIOR FILING DATE: 1999-10-08  
; PRIOR APPLICATION NUMBER: USSN 60/103,794  
; PRIOR FILING DATE: 1998-10-09  
; PRIOR APPLICATION NUMBER: USSN 60/132,068  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: PCT/US99/25373  
; PRIOR FILING DATE: 1999-10-08  
; NUMBER OF SEQ ID NOS: 1068  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 760  
; LENGTH: 445  
; TYPE: DNA  
; ORGANISM: Neisseria meningitidis  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (286)..(286)  
; OTHER INFORMATION: N is any nucleotide  
US-10-915-740A-760

Query Match 84.2%; Score 16; DB 9; Length 445;  
Best Local Similarity 100.0%; Pred. No. 4.8e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 CAGTGCATCCCAATTG 18  
|||||  
DB 263 CAGTGCATCCCAATTG 248

RESULT 22  
US-10-972-079-87794/c  
; Sequence 87794, Application US/10972079  
; Publication No. US20050153317A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: ROSENFELD, David  
; APPLICANT: KERR, Richard  
; APPLICANT: BATES, Stephen  
; APPLICANT: HOLM, Tom  
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEE-  
; TITLE OF INVENTION: LIVESTOCK  
; FILE REFERENCE: MM11110-2  
; CURRENT APPLICATION NUMBER: US/10/972,079  
; CURRENT FILING DATE: 2004-10-22  
; PRIOR APPLICATION NUMBER: US 60/514,333  
; PRIOR FILING DATE: 2003-10-24  
; NUMBER OF SEQ ID NOS: 96631  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 87794  
; LENGTH: 564  
; TYPE: DNA  
; ORGANISM: Chicken 19866894390100\_8  
US-10-972-079-87794

Query Match 84.2%; Score 16; DB 9; Length 564;  
Best Local Similarity 88.9%; Pred. No. 4.9e+02;  
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCCAATTG 18  
|||||  
DB 189 CCCAGTGCATCCCAATTG 172

RESULT 23  
US-09-925-065A-248192/c  
; Sequence 248192, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; FILE REFERENCE: 108927.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; PRIOR FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 248192  
; LENGTH: 589  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-248192

Query Match 84.2%; Score 16; DB 4; Length 589;  
Best Local Similarity 100.0%; Pred. No. 4.9e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 CAGTGCATCCAAATTG 18  
|||||:|||||  
Db 303 CAGTGCATCCAAATTG 288

RESULT 24  
US-09-925-065A-248193/c  
; Sequence 248193, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; FILE REFERENCE: 108927.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; PRIOR FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 248193  
; LENGTH: 589  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-248193

Query Match 84.2%; Score 16; DB 4; Length 589;  
Best Local Similarity 100.0%; Pred. No. 4.9e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 CAGTGCATCCAAATTG 18  
|||||:|||||

Db 303 CAGTGCATCCAAATTG 288  
RESULT 25  
US-10-972-079-87793/c  
; Sequence 87793, Application US/10972079  
; Publication No. US20050153317A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: ROSENFELD, David  
; APPLICANT: KERR, Richard  
; APPLICANT: BATES, Stephen  
; APPLICANT: HOLM, Tom  
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BREED  
; FILE REFERENCE: MM11110-2  
; CURRENT APPLICATION NUMBER: US/10/972,079  
; CURRENT FILING DATE: 2004-10-22  
; PRIOR APPLICATION NUMBER: US 60/514,333  
; PRIOR FILING DATE: 2003-10-24  
; NUMBER OF SEQ ID NOS: 96631  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 87793  
; LENGTH: 592  
; TYPE: DNA  
; ORGANISM: Chicken 19866894390100\_7  
US-10-972-079-87793

Query Match 84.2%; Score 16; DB 9; Length 592;  
Best Local Similarity 88.9%; Pred. No. 4.9e+02;  
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATTG 18  
|||||:|||||  
Db 217 CCCAGTGCATCCAAATTG 200

RESULT 26  
US-10-972-079-87791/c  
; Sequence 87791, Application US/10972079  
; Publication No. US20050153317A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: ROSENFELD, David  
; APPLICANT: KERR, Richard  
; APPLICANT: BATES, Stephen  
; APPLICANT: HOLM, Tom  
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BREED  
; FILE REFERENCE: MM11110-2  
; CURRENT APPLICATION NUMBER: US/10/972,079  
; CURRENT FILING DATE: 2004-10-22  
; PRIOR APPLICATION NUMBER: US 60/514,333  
; PRIOR FILING DATE: 2003-10-24  
; NUMBER OF SEQ ID NOS: 96631  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 87791  
; LENGTH: 600  
; TYPE: DNA  
; ORGANISM: Chicken 19866894390100\_5  
US-10-972-079-87791

Query Match 84.2%; Score 16; DB 9; Length 600;  
Best Local Similarity 88.9%; Pred. No. 4.9e+02;  
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATTG 18  
|||||:|||||  
Db 306 CCCAGTGCATCCAAATTG 289

```
RESULT 27
US-10-972-079-87792/c
; Sequence 87792, Application US/10972079
; Publication No. US20050153317A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE Sue K.
; APPLICANT: ROSENFELD, David
; APPLICANT: KERR, Richard
; APPLICANT: BATES, Stephen
; APPLICANT: HOLM, Tom
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BREED
; TITLE OF INVENTION: LIVESTOCK
; FILE REFERENCE: MM1110-2
; CURRENT APPLICATION NUMBER: US/10/972,079
; CURRENT FILING DATE: 2004-10-22
; PRIOR APPLICATION NUMBER: US 60/514,333
; PRIOR FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 96631
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 87792
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Chicken 19866894390100_6
US-10-972-079-87792

Query Match      84.2%; Score 16; DB 9; Length 600;
Best Local Similarity 88.9%; Pred. No. 4.9e+02;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATTG 18
DB 280 CCCAGTGCATCCAAATTG 263

RESULT 28
US-09-864-761-7436/c
; Sequence 7436, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
```

```

; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 7436
; LENGTH: 512
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC019194.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2
US-09-864-761-7436

Query Match      83.2%; Score 15.8; DB 3; Length 512;
Best Local Similarity 89.5%; Pred. No. 6.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATTGA 19
DB 318 CCCAGGGCTTCCAAATTGA 300

RESULT 29
US-09-925-065A-190164
; Sequence 190164, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 190164
; LENGTH: 529
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-190164

Query Match      83.2%; Score 15.8; DB 4; Length 529;
Best Local Similarity 89.5%; Pred. No. 6.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATTGA 19
```

Db 142 CCCATTTCATCCAAATTGA 160

RESULT 30  
US-10-972-079-73299/c  
; Sequence 73299, Application US/10972079  
; Publication No. US20050153317A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: ROSENFELD, David  
; APPLICANT: KERR, Richard  
; APPLICANT: BATES, Stephen  
; APPLICANT: HOLM, Tom  
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEER  
; TITLE OF INVENTION: LIVESTOCK  
; FILE REFERENCE: MM1110-2  
; CURRENT APPLICATION NUMBER: US/10/972,079  
; CURRENT FILING DATE: 2004-10-22  
; PRIOR APPLICATION NUMBER: US 60/514,333  
; PRIOR FILING DATE: 2003-10-24  
; NUMBER OF SEQ ID NOS: 96631  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 73299  
; LENGTH: 582  
; TYPE: DNA  
; ORGANISM: Chicken 19866894354273\_3  
US-10-972-079-73299

Query Match 83.2%; Score 15.8; DB 9; Length 582;  
Best Local Similarity 89.5%; Pred. No. 6.2e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATTGA 19  
Db 517 CACAGTGCATCCAAATGA 499

RESULT 31  
US-09-925-065A-918407  
; Sequence 918407, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 918407  
; LENGTH: 586  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-918407

Query Match 83.2%; Score 15.8; DB 4; Length 586;  
Best Local Similarity 89.5%; Pred. No. 6.2e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATTGA 19

Db 537 CCCAGTGCATCTGAATTGA 555

RESULT 32  
US-10-972-079-25977  
; Sequence 25977, Application US/10972079  
; Publication No. US20050153317A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: ROSENFELD, David  
; APPLICANT: KERR, Richard  
; APPLICANT: BATES, Stephen  
; APPLICANT: HOLM, Tom  
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEER  
; TITLE OF INVENTION: LIVESTOCK  
; FILE REFERENCE: MM1110-2  
; CURRENT APPLICATION NUMBER: US/10/972,079  
; CURRENT FILING DATE: 2004-10-22  
; PRIOR APPLICATION NUMBER: US 60/514,333  
; PRIOR FILING DATE: 2003-10-24  
; NUMBER OF SEQ ID NOS: 96631  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 25977  
; LENGTH: 599  
; TYPE: DNA  
; ORGANISM: Chicken 19866894241792\_1  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(6)  
; OTHER INFORMATION: n is any nucleotide  
US-10-972-079-25977

Query Match 83.2%; Score 15.8; DB 9; Length 599;  
Best Local Similarity 89.5%; Pred. No. 6.3e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATTGA 19  
Db 36 CCCAGAGCATCCAAAATGA 54

RESULT 33  
US-10-972-079-25978  
; Sequence 25978, Application US/10972079  
; Publication No. US20050153317A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: ROSENFELD, David  
; APPLICANT: KERR, Richard  
; APPLICANT: BATES, Stephen  
; APPLICANT: HOLM, Tom  
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEER  
; TITLE OF INVENTION: LIVESTOCK  
; FILE REFERENCE: MM1110-2  
; CURRENT APPLICATION NUMBER: US/10/972,079  
; CURRENT FILING DATE: 2004-10-22  
; PRIOR APPLICATION NUMBER: US 60/514,333  
; PRIOR FILING DATE: 2003-10-24  
; NUMBER OF SEQ ID NOS: 96631  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 25978  
; LENGTH: 600  
; TYPE: DNA  
; ORGANISM: Chicken 19866894241792\_2  
US-10-972-079-25978

Query Match 83.2%; Score 15.8; DB 9; Length 600;  
Best Local Similarity 89.5%; Pred. No. 6.3e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATTGA 19

Db 27 CCCAGAGCATCCAAATGA 45  
||||| ||||||| |||

## RESULT 34

US-10-972-079-25979  
; Sequence 25979, Application US/10972079  
; Publication No. US2005015331/1A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: ROSENFELD, David  
; APPLICANT: KERR, Richard  
; APPLICANT: BATES, Stephen  
; APPLICANT: HOLM, Tom  
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEER  
; FILE REFERENCE: WM1110-2  
; CURRENT APPLICATION NUMBER: US/10/972,079  
; PRIOR FILING DATE: 2004-10-22  
; PRIOR APPLICATION NUMBER: US 60/514,333  
; PRIOR FILING DATE: 2003-10-24  
; NUMBER OF SEQ ID NOS: 96631  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 25979  
; LENGTH: 600  
; TYPE: DNA  
; ORGANISM: Chicken 19866894241792\_3  
US-10-972-079-25979

Query Match 83.2%; Score 15.8; DB 9; Length 600;  
Best Local Similarity 89.5%; Pred. No. 6.3e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATGA 19  
||||| ||||||| |||

Db 8 CCCAGAGCATCCAAATGA 26

## RESULT 35

US-10-972-079-73298/c  
; Sequence 73298, Application US/10972079  
; Publication No. US2005015331/1A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: ROSENFELD, David  
; APPLICANT: KERR, Richard  
; APPLICANT: BATES, Stephen  
; APPLICANT: HOLM, Tom  
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEER  
; FILE REFERENCE: WM1110-2  
; CURRENT APPLICATION NUMBER: US/10/972,079  
; PRIOR FILING DATE: 2004-10-22  
; PRIOR APPLICATION NUMBER: US 60/514,333  
; PRIOR FILING DATE: 2003-10-24  
; NUMBER OF SEQ ID NOS: 96631  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 73298  
; LENGTH: 600  
; TYPE: DNA  
; ORGANISM: Chicken 19866894354273\_2  
US-10-972-079-73298

Query Match 83.2%; Score 15.8; DB 9; Length 600;  
Best Local Similarity 89.5%; Pred. No. 6.3e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATGA 19  
||||| ||||||| |||

Db 570 CACAGTGCATCCAAATGA 552

## RESULT 36

US-10-960-409-5  
; Sequence 5, Application US/10960409  
; Publication No. US20050089912A1  
; GENERAL INFORMATION:  
; APPLICANT: Armour, Christopher D.  
; APPLICANT: Castle, John C.  
; APPLICANT: Garrett-Englele, Philip W.  
; APPLICANT: Kan, Zhengyan  
; APPLICANT: Raymond, Christopher K.  
; APPLICANT: Tsinoemas, Nicholas F.  
; TITLE OF INVENTION: ALTERNATIVELY SPLICED ISOFORMS OF  
; FILE REFERENCE: RS0216  
; CURRENT APPLICATION NUMBER: US/10/960,409  
; PRIOR FILING DATE: 2004-10-06  
; PRIOR APPLICATION NUMBER: US 60/509,361  
; PRIOR FILING DATE: 2003-10-07  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 2538  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-960-409-5

Query Match 83.2%; Score 15.8; DB 9; Length 2538;  
Best Local Similarity 89.5%; Pred. No. 7.2e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATTGA 19  
||||| ||||||| |||

Db 736 CCCAATGCATCCCAACTTGA 754

## RESULT 37

US-10-960-409-3  
; Sequence 3, Application US/10960409  
; Publication No. US20050089912A1  
; GENERAL INFORMATION:  
; APPLICANT: Armour, Christopher D.  
; APPLICANT: Castle, John C.  
; APPLICANT: Garrett-Englele, Philip W.  
; APPLICANT: Kan, Zhengyan  
; APPLICANT: Raymond, Christopher K.  
; APPLICANT: Tsinoemas, Nicholas F.  
; TITLE OF INVENTION: ALTERNATIVELY SPLICED ISOFORMS OF  
; FILE REFERENCE: RS0216  
; CURRENT APPLICATION NUMBER: US/10/960,409  
; PRIOR FILING DATE: 2004-10-06  
; PRIOR APPLICATION NUMBER: US 60/509,361  
; PRIOR FILING DATE: 2003-10-07  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 2811  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-960-409-3

Query Match 83.2%; Score 15.8; DB 9; Length 2811;  
Best Local Similarity 89.5%; Pred. No. 7.3e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATTGA 19  
||||| ||||||| |||

Db 637 CCCAATGCATCCCAACTTGA 655

## RESULT 38

US-10-394-948-30

```
; Sequence 30, Application US/10394948
; Publication No. US20040023267A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; TITLE OF INVENTION: No. US20040023267A1el Compositions and Methods in Cancer
; FILE REFERENCE: 529452000900
; CURRENT APPLICATION NUMBER: US/10/394,948
; CURRENT FILING DATE: 2003-03-21
; PRIOR APPLICATION NUMBER: US 60/367,025
; PRIOR FILING DATE: 2002-03-21
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 2907
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-394-948-30

Query Match      83.2%; Score 15.8; DB 7; Length 2907;
Best Local Similarity 89.5%; Pred. No. 7.3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATTGA 19
||||| ||||| ||||| |||||
Db 733 CCCATGTCATCCCAACTTGA 751

RESULT 39
US-10-052-482-120
; Sequence 120, Application US/10052482
; Publication No. US20040072264A1
; GENERAL INFORMATION:
; APPLICANT: Engelhard, Eric
; APPLICANT: Morris, David
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: A-71087/RMS/DCF
; CURRENT APPLICATION NUMBER: US/10/052,482
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 241
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 120
; LENGTH: 2907
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-052-482-120

Query Match      83.2%; Score 15.8; DB 7; Length 2907;
Best Local Similarity 89.5%; Pred. No. 7.3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCCAATTGA 19
||||| ||||| ||||| |||||
Db 733 CCCATGTCATCCCAACTTGA 751

RESULT 40
US-10-247-671-22
; Sequence 22, Application US/10247671
; Publication No. US20030194721A1
; GENERAL INFORMATION:
; APPLICANT: Mikita, Thomas
; APPLICANT: Shiffman, Dov
; APPLICANT: Porter, Gordon, J.
; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED FOAM CELLS
; FILE REFERENCE: PA-0050 US
; CURRENT APPLICATION NUMBER: US/10/247,671
; CURRENT FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: 60/323,784
```

```
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 186
; SOFTWARE: PERL Program
; SEQ ID NO 22
; LENGTH: 3497
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20030194721A1 1438184CB1
US-10-247-671-22

Query Match      83.2%; Score 15.8; DB 6; Length 3497;
Best Local Similarity 89.5%; Pred. No. 7.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCCAATTGA 19
||||| ||||| ||||| |||||
Db 826 CCCAATGTCATCCCAACTTGA 844

Search completed: November 21, 2005, 05:03:42
Job time : 573.346 secs
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2005, 21:55:46 ; Search time 265.531 Seconds  
(without alignments)  
9.405 Million cell updates/sec

Title: US-10-627-757-27

Perfect score: 19

Sequence: 1 cccagtgcattccaaattga 19

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3196817 seqs, 65720914 residues

Total number of hits satisfying chosen parameters: 6393634

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA New.\*

- 1: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
- 2: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
- 5: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*
- 7: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq2.\*
- 9: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq3.\*
- 10: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	15.4	81.1	19	8	US-11-101-244-119060
2	15.4	81.1	19	9	US-11-083-784-119060
3	14.4	75.8	19	8	US-11-101-244-1448592
4	14.4	75.8	19	8	US-11-101-244-1448633
5	14.4	75.8	19	8	US-11-101-244-1448650
6	14.4	75.8	19	9	US-11-083-784-1448592
7	14.4	75.8	19	9	US-11-083-784-1448633
8	14.4	75.8	19	9	US-11-083-784-1448650
9	14.2	74.7	4431	1	US-10-847-956A-3
10	14	73.7	19	8	US-11-101-244-381869
11	14	73.7	19	8	US-11-101-244-381966
12	14	73.7	19	8	US-11-101-244-382061
13	14	73.7	19	9	US-11-083-784-381869
14	14	73.7	19	9	US-11-083-784-381966
15	14	73.7	19	9	US-11-083-784-382061
16	13.8	72.6	19	8	US-11-101-244-335695
17	13.8	72.6	19	8	US-11-101-244-335695
18	13.8	72.6	19	8	US-11-101-244-1001268
19	13.8	72.6	19	8	US-11-101-244-1001280
20	13.8	72.6	19	8	US-11-101-244-1001293
21	13.8	72.6	19	8	US-11-101-244-1402319
22	13.8	72.6	19	8	US-11-101-244-1402328
23	13.8	72.6	19	9	US-11-083-784-246861

ALIGNMENTS

RESULT 1

US-11-101-244-119060  
; Sequence 119060, Application US/1101244  
; Publication No. US20050246794A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmacon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 134990US  
; CURRENT APPLICATION NUMBER: US/11/101,244  
; CURRENT FILING DATE: 2005-04-07  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 119060  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-101-244-119060

Query Match 81.1%; Score 15.4; DB 8; Length 19;  
Best Local Similarity 70.6%; Pred. No. 43;  
Matches 12; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATT 17

Db 1 CCCAUGCAUCCAAAUU 17

RESULT 2

US-11-083-784-119060  
; Sequence 119060, Application US/11083784  
; Publication No. US20050245475A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmacon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William

Sequence 335695,  
Sequence 1001268,  
Sequence 1001280,  
Sequence 1001293,  
Sequence 1402319,  
Sequence 1402328,  
Sequence 223, App  
Sequence 246868,  
Sequence 1025272,  
Sequence 1244732,  
Sequence 1244785,  
Sequence 246868,  
Sequence 1025272,  
Sequence 1244732,  
Sequence 1244785,  
Sequence 3, Appli  
Sequence 461849,  
Sequence 578651,  
Sequence 1259900,  
Sequence 1260288,  
Sequence 1367244,  
Sequence 1372931,

```
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR APPLICATION NUMBER: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 119060
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-119060
```

```
Query Match      81.1%; Score 15.4; DB 9; Length 19;
Best Local Similarity 70.8%; Pred. No. 43;
Matches 12; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATTG 17
    ||||:|||||:
Db 1 CCCAUGCAUCCAAAUU 17
```

```
RESULT 3
US-11-101-244-1448592
; Sequence 1448592, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 1448592
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-1448592
```

```
Query Match      75.8%; Score 14.4; DB 8; Length 19;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 12; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 4 AGTGCATCCAAATTGA 19
    ||||:|||||:
Db 4 AGUGCAUCCAAAGUGA 19
```

```
RESULT 4
US-11-101-244-1448633
; Sequence 1448633, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
```

```
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 1448633
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-1448633
```

```
Query Match      75.8%; Score 14.4; DB 8; Length 19;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 12; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 4 AGTGCATCCAAATTGA 19
    ||||:|||||:
Db 2 AGUGCAUCCAAAGUGA 17
```

```
RESULT 5
US-11-101-244-1448650
; Sequence 1448650, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 1448650
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-1448650
```

```
Query Match      75.8%; Score 14.4; DB 8; Length 19;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 12; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 4 AGTGCATCCAAATTGA 19
    ||||:|||||:
Db 3 AGUGCAUCCAAAGUGA 18
```

```
RESULT 6
US-11-083-784-1448592
; Sequence 1448592, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
```

```
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR APPLICATION NUMBER: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 1448592
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-1448592
```

```
Query Match 75.8%; Score 14.4; DB 9; Length 19;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 12; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 4 AGTGCATCCAAATTGA 19
||:||||:||||:|
DB 4 AGUGCAUCCAAAGUGA 19
```

```
RESULT 7
US-11-083-784-1448633
; Sequence 1448633, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR APPLICATION NUMBER: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 1448633
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-1448633
```

```
Query Match 75.8%; Score 14.4; DB 9; Length 19;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 12; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 4 AGTGCATCCAAATTGA 19
||:||||:||||:|
DB 2 AGUGCAUCCAAAGUGA 17
```

```
RESULT 8
US-11-083-784-1448650
; Sequence 1448650, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
```

```
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR APPLICATION NUMBER: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 1448650
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-1448650
```

```
Query Match 75.8%; Score 14.4; DB 9; Length 19;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 12; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 4 AGTGCATCCAAATTGA 19
||:||||:||||:|
DB 3 AGUGCAUCCAAAGUGA 18
```

```
RESULT 9
US-10-647-956A-3/c
; Sequence 3, Application US/10647956A
; Publication No. US20050251878A1
; GENERAL INFORMATION:
; APPLICANT: French-Constant, Richard
; APPLICANT: Bowen, David
; APPLICANT: Rocheleau, Thomas
; APPLICANT: Waterfield, Nicholas
; TITLE OF INVENTION: DNA SEQUENCES FROM PHOTORHABDUS LUMINESCENS
; FILE REFERENCE: 61845
; CURRENT APPLICATION NUMBER: US/10/647,956A
; CURRENT FILING DATE: 2003-08-26
; PRIOR APPLICATION NUMBER: CURRENT APPLICATION NUMBER: US/09/817,514
; PRIOR FILING DATE: CURRENT FILING DATE: 2000-03-26
; PRIOR APPLICATION NUMBER: US 60/191806
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 4431
; TYPE: DNA
; ORGANISM: Photorhabdus luminescens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(4431)
US-10-647-956A-3
```

```
Query Match 74.7%; Score 14.2; DB 1; Length 4431;
Best Local Similarity 84.2%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 CCCAGTGCATCCAAATTGA 19
|||||:|||||:|
DB 3632 CCCAGATCCCAAGGTGA 3614
```

```
RESULT 10
US-11-101-244-381869
; Sequence 381869, Application US/1101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
```

; APPLICANT: Dharmacon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/101,244  
; CURRENT FILING DATE: 2005-04-07  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 381869  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-101-244-381869

Query Match 73.7%; Score 14; DB 8; Length 19;  
Best Local Similarity 85.7%; Pred. No. 1.9e+02;  
Matches 12; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAA 14  
|||||:|||||  
Db 1 CCCAGUGCAUCCAA 14

RESULT 11  
US-11-101-244-381966  
; Sequence 381966, Application US/11101244  
; Publication No. US20050246794A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmacon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/101,244  
; CURRENT FILING DATE: 2005-04-07  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 381966  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-101-244-381966

Query Match 73.7%; Score 14; DB 8; Length 19;  
Best Local Similarity 85.7%; Pred. No. 1.9e+02;  
Matches 12; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAA 14  
|||||:|||||  
Db 1 CCCAGUGCAUCCAA 14

RESULT 12  
US-11-101-244-382061  
; Sequence 382061, Application US/11101244  
; Publication No. US20050246794A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmacon, Inc.

; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/101,244  
; CURRENT FILING DATE: 2005-04-07  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 382061  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-101-244-382061

Query Match 73.7%; Score 14; DB 8; Length 19;  
Best Local Similarity 85.7%; Pred. No. 1.9e+02;  
Matches 12; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAA 14  
|||||:|||||  
Db 1 CCCAGUGCAUCCAA 14

RESULT 13  
US-11-083-784-381869  
; Sequence 381869, Application US/11083784  
; Publication No. US20050245475A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmacon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/083,784  
; CURRENT FILING DATE: 2005-03-18  
; PRIOR APPLICATION NUMBER: US/10/714,333  
; PRIOR FILING DATE: 2003-11-14  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 381869  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-083-784-381869

Query Match 73.7%; Score 14; DB 9; Length 19;  
Best Local Similarity 85.7%; Pred. No. 1.9e+02;  
Matches 12; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAA 14  
|||||:|||||  
Db 1 CCCAGUGCAUCCAA 14

RESULT 14  
US-11-083-784-381966  
; Sequence 381966, Application US/11083784  
; Publication No. US20050245475A1  
; GENERAL INFORMATION:

; APPLICANT: Dharmacon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/083,784  
; CURRENT FILING DATE: 2005-03-18  
; PRIOR FILING DATE: US/10/714,333  
; PRIOR FILING DATE: 2003-11-14  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 381966  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-083-784-381966

Query Match 73.7%; Score 14; DB 9; Length 19;  
Best Local Similarity 85.7%; Pred. No. 1.9e+02;  
Matches 12; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAA 14  
| | | | | : | | | | |  
Db 1 CCCAGUGCAUCCAA 14

## RESULT 15

US-11-083-784-382061  
; Sequence 382061, Application US/11083784  
; Publication No. US20050245475A1  
; GENERAL INFORMATION:

; APPLICANT: Dharmacon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/083,784  
; CURRENT FILING DATE: 2005-03-18  
; PRIOR FILING DATE: US/10/714,333  
; PRIOR FILING DATE: 2003-11-14  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 382061  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-083-784-382061

Query Match 73.7%; Score 14; DB 9; Length 19;  
Best Local Similarity 85.7%; Pred. No. 1.9e+02;  
Matches 12; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAA 14  
| | | | | : | | | | |  
Db 1 CCCAGUGCAUCCAA 14

## RESULT 16

US-11-101-244-246861

; Sequence 246861, Application US/11101244  
; Publication No. US20050246794A1  
; GENERAL INFORMATION:

; APPLICANT: Dharmacon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/101,244  
; CURRENT FILING DATE: 2005-04-07  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 246861  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-101-244-246861

Query Match 72.6%; Score 13.8; DB 8; Length 19;  
Best Local Similarity 70.6%; Pred. No. 2.3e+02;  
Matches 12; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATT 17  
| | | | | : | | | | |  
Db 2 CGCAGUGCAGCCAAAUU 18

## RESULT 17

US-11-101-244-335695  
; Sequence 335695, Application US/11101244  
; Publication No. US20050246794A1  
; GENERAL INFORMATION:

; APPLICANT: Dharmacon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/101,244  
; CURRENT FILING DATE: 2005-04-07  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 335695  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-101-244-335695

Query Match 72.6%; Score 13.8; DB 8; Length 19;  
Best Local Similarity 64.7%; Pred. No. 2.3e+02;  
Matches 11; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 3 CAGTGCATCCAAATTGA 19  
| | | | | : | | | | |  
Db 1 CAAUGCAUCCAACTUGA 17

## RESULT 18

US-11-101-244-1001268  
; Sequence 1001268, Application US/11101244



```
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 1402328
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-1402328
```

```
Query Match 72.6%; Score 13.8; DB 8; Length 19;
Best Local Similarity 64.7%; Pred. No. 2.3e+02;
Matches 11; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 1 CCCAGTGCATCCAAATT 17
|||:|||||:
Db 1 CACAGUCCAUCCAAAUU 17
```

## RESULT 23

```
US-11-083-784-246861
; Sequence 246861, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR APPLICATION NUMBER: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 246861
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-246861
```

```
Query Match 72.6%; Score 13.8; DB 9; Length 19;
Best Local Similarity 70.6%; Pred. No. 2.3e+02;
Matches 12; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 1 CCCAGTGCATCCAAATT 17
|||:|||||:
Db 2 CGCAGUGCAGCCAAAUU 18
```

## RESULT 24

```
US-11-083-784-335695
; Sequence 335695, Application US/11083784
; Publication No. US20050245475A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR APPLICATION NUMBER: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 335695
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-335695
```

```
Query Match 72.6%; Score 13.8; DB 9; Length 19;
Best Local Similarity 64.7%; Pred. No. 2.3e+02;
Matches 11; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 3 CAGTGCATCCAAATTGA 19
|||:|||||:
Db 1 CAAUGCAUCCAAAUUGA 17
```

## RESULT 25

```
US-11-083-784-1001268
; Sequence 1001268, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR APPLICATION NUMBER: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 1001268
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-1001268
```

```
Query Match 72.6%; Score 13.8; DB 9; Length 19;
Best Local Similarity 64.7%; Pred. No. 2.3e+02;
Matches 11; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 1 CCCAGTGCATCCAAATT 17
|||:|||||:
Db 2 CCAAGUGUCCAAAUU 18
```

## RESULT 26

US-11-083-784-1001280  
; Sequence 1001280, Application US/11083784  
; Publication No. US20050245475A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmacon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/083,784  
; CURRENT FILING DATE: 2005-03-18  
; PRIOR APPLICATION NUMBER: US/10/714,333  
; PRIOR FILING DATE: 2003-11-14  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 1001280  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-083-784-1001280

Query Match 72.6%; Score 13.8; DB 9; Length 19;  
Best Local Similarity 64.7%; Pred. No. 2.3e+02;  
Matches 11; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATT 17  
|||:|||||:  
Db 1 CCAAGUGCUCCAAAUU 17

RESULT 27  
US-11-083-784-1001293  
; Sequence 1001293, Application US/11083784  
; Publication No. US20050245475A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmacon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/083,784  
; CURRENT FILING DATE: 2005-03-18  
; PRIOR APPLICATION NUMBER: US/10/714,333  
; PRIOR FILING DATE: 2003-11-14  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 1001293  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-083-784-1001293

Query Match 72.6%; Score 13.8; DB 9; Length 19;  
Best Local Similarity 64.7%; Pred. No. 2.3e+02;  
Matches 11; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATT 17  
|||:|||||:  
Db 3 CCAAGUGCUCCAAAUU 19

RESULT 28  
US-11-083-784-1402319  
; Sequence 1402319, Application US/11083784  
; Publication No. US20050245475A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmacon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/083,784  
; CURRENT FILING DATE: 2005-03-18  
; PRIOR APPLICATION NUMBER: US/10/714,333  
; PRIOR FILING DATE: 2003-11-14  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 1402319  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-083-784-1402319

Query Match 72.6%; Score 13.8; DB 9; Length 19;  
Best Local Similarity 64.7%; Pred. No. 2.3e+02;  
Matches 11; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCCAGTGCATCCAAATT 17  
|||:|||||:  
Db 2 CACAGUCCAUCCAAAUU 18

RESULT 29  
US-11-083-784-1402328  
; Sequence 1402328, Application US/11083784  
; Publication No. US20050245475A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmacon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/083,784  
; CURRENT FILING DATE: 2005-03-18  
; PRIOR APPLICATION NUMBER: US/10/714,333  
; PRIOR FILING DATE: 2003-11-14  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 1402328  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-083-784-1402328

Query Match 72.6%; Score 13.8; DB 9; Length 19;  
Best Local Similarity 64.7%; Pred. No. 2.3e+02;  
Matches 11; Conservative 4; Mismatches 2; Indels 0; Gaps 0;





```
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 1244732
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-1244732
```

```
Query Match 70.5%; Score 13.4; DB 8; Length 19;
Best Local Similarity 73.3%; Pred. No. 3.5e+02;
Matches 11; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 2 CCAGTGCATCCCAAT 16
   |||:||||:||||:
Db 3 CCAAGCAUCCAAAU 17
```

```
RESULT 34
US-11-101-244-1244785
; Sequence 1244785, Application US/1101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 1244785
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-1244785
```

```
Query Match 70.5%; Score 13.4; DB 8; Length 19;
Best Local Similarity 73.3%; Pred. No. 3.5e+02;
Matches 11; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 2 CCAGTGCATCCCAAT 16
   |||:||||:||||:
Db 2 CCAAGCAUCCAAAU 16
```

```
RESULT 35
US-11-083-784-246868
; Sequence 246868, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
```

```
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR APPLICATION NUMBER: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 246868
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-246868
```

```
Query Match 70.5%; Score 13.4; DB 9; Length 19;
Best Local Similarity 73.3%; Pred. No. 3.5e+02;
Matches 11; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 3 CAGTGCATCCCAATT 17
   |||:||||:||||:
Db 2 CAGUGCAGCCAAAU 16
```

```
RESULT 36
US-11-083-784-1025272/c
; Sequence 1025272, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR APPLICATION NUMBER: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 1025272
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-1025272
```

```
Query Match 70.5%; Score 13.4; DB 9; Length 19;
Best Local Similarity 93.3%; Pred. No. 3.5e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 4 AGTGCATCCCAATTG 18
   |||:||||:||||:
Db 18 AGTTCATCCCAATTG 4
```

```
RESULT 37
US-11-083-784-1244732
; Sequence 1244732, Application US/11083784
```

Publication No. US20050245475A1  
GENERAL INFORMATION:  
APPLICANT: Dharmacon, Inc.  
APPLICANT: Khvorova, Anastasia  
APPLICANT: Reynolds, Angela  
APPLICANT: Leake, Devin  
APPLICANT: Marshall, William  
APPLICANT: Scaringe, Stephen  
TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
FILE REFERENCE: 13499US  
CURRENT APPLICATION NUMBER: US/11/083,784  
CURRENT FILING DATE: 2005-03-18  
PRIOR APPLICATION NUMBER: US/10/714,333  
PRIOR FILING DATE: 2003-11-14  
PRIOR APPLICATION NUMBER: 60/502,050  
PRIOR FILING DATE: 2003-09-10  
PRIOR APPLICATION NUMBER: 60/426,137  
PRIOR FILING DATE: 2002-11-14  
NUMBER OF SEQ ID NOS: 1591911  
SOFTWARE: Proprietary  
SEQ ID NO 1244732  
LENGTH: 19  
TYPE: RNA  
ORGANISM: Homo sapiens  
US-11-083-784-1244732

Query Match 70.5%; Score 13.4; DB 9; Length 19;  
Best Local Similarity 73.3%; Pred. No. 3.5e+02;  
Matches 11; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCAGTGCATCCAAAT 16  
|||:||||:||||:  
Db 3 CCAAUCAUCCAAAU 17

RESULT 38  
US-11-083-784-1244785  
Sequence 1244785, Application US/11083784  
Publication No. US20050245475A1  
GENERAL INFORMATION:  
APPLICANT: Dharmacon, Inc.  
APPLICANT: Khvorova, Anastasia  
APPLICANT: Reynolds, Angela  
APPLICANT: Leake, Devin  
APPLICANT: Marshall, William  
APPLICANT: Scaringe, Stephen  
TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
FILE REFERENCE: 13499US  
CURRENT APPLICATION NUMBER: US/11/083,784  
CURRENT FILING DATE: 2005-03-18  
PRIOR APPLICATION NUMBER: US/10/714,333  
PRIOR FILING DATE: 2003-11-14  
PRIOR APPLICATION NUMBER: 60/502,050  
PRIOR FILING DATE: 2003-09-10  
PRIOR APPLICATION NUMBER: 60/426,137  
PRIOR FILING DATE: 2002-11-14  
NUMBER OF SEQ ID NOS: 1591911  
SOFTWARE: Proprietary  
SEQ ID NO 1244785  
LENGTH: 19  
TYPE: RNA  
ORGANISM: Homo sapiens  
US-11-083-784-1244785

Query Match 70.5%; Score 13.4; DB 9; Length 19;  
Best Local Similarity 73.3%; Pred. No. 3.5e+02;  
Matches 11; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCAGTGCATCCAAAT 16  
|||:||||:||||:  
Db 2 CCAAUCAUCCAAAU 16

RESULT 39  
US-11-102-978-3  
Sequence 3, Application US/11102978  
Publication No. US20050250142A1  
GENERAL INFORMATION:  
APPLICANT: University of Utah Technology Transfer Office  
APPLICANT: University of Utah Research Foundation  
TITLE OF INVENTION: Diagnosis and Treatment of Herpes Simplex Virus Disease  
FILE REFERENCE: 0274-5537.1US  
CURRENT APPLICATION NUMBER: US/11/102,978  
CURRENT FILING DATE: 2005-04-11  
PRIOR APPLICATION NUMBER: PCT/US2003/033152  
PRIOR FILING DATE: 2003-10-18  
PRIOR APPLICATION NUMBER: 60/419,576  
PRIOR FILING DATE: 2002-10-18  
NUMBER OF SEQ ID NOS: 13  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 3  
LENGTH: 340000  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: exon  
LOCATION: (56948)..(57115)  
OTHER INFORMATION: C21orf34 exon  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (80056)..(81089)  
OTHER INFORMATION: Gene VDACP; voltage-dependent anion channel isoform 2 pseudogen  
FEATURE:  
NAME/KEY: exon  
LOCATION: (167308)..(167438)  
OTHER INFORMATION: C21orf34 exon  
FEATURE:  
NAME/KEY: exon  
LOCATION: (216732)..(216833)  
OTHER INFORMATION: C21orf34 exon  
US-11-102-978-3

Query Match 70.5%; Score 13.4; DB 7; Length 340000;  
Best Local Similarity 93.3%; Pred. No. 2.1e+02;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCAGTGCATCCAAAT 16  
|||||:|||||:|||||:  
Db 19321 CCAGTGCATCCAAAT 19335

RESULT 40  
US-11-101-244-461849  
Sequence 461849, Application US/11101244  
Publication No. US20050246794A1  
GENERAL INFORMATION:  
APPLICANT: Dharmacon, Inc.  
APPLICANT: Khvorova, Anastasia  
APPLICANT: Reynolds, Angela  
APPLICANT: Leake, Devin  
APPLICANT: Marshall, William  
APPLICANT: Scaringe, Stephen  
TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
FILE REFERENCE: 13499US  
CURRENT APPLICATION NUMBER: US/11/101,244  
CURRENT FILING DATE: 2005-04-07  
PRIOR APPLICATION NUMBER: 60/502,050  
PRIOR FILING DATE: 2003-09-10  
PRIOR APPLICATION NUMBER: 60/426,137  
PRIOR FILING DATE: 2002-11-14  
NUMBER OF SEQ ID NOS: 1591911  
SOFTWARE: Proprietary  
SEQ ID NO 461849  
LENGTH: 19  
TYPE: RNA  
ORGANISM: Homo sapiens

US-11-101-244-461849

Query Match 69.5%; Score 13.2; DB 8; Length 19;  
Best Local Similarity 61.1%; Pred. No. 4.3e+02;  
Matches 11; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY	2	CCAGTGCATCCAAATTGA	19
DB	2	CCAGUACAUCCAGUUUA	19

Search completed: November 21, 2005, 05:22:42  
Job time : 266.531 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2005, 16:48:26 ; Search time 805.309 Seconds  
(without alignments)  
1552.890 Million cell updates/sec

Title: US-10-627-757-28

Perfect score: 22  
Sequence: 1 tcattgtcacattactgga 22

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues  
Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_in.\*  
3: gb\_env.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pr.\*  
9: gb\_ro.\*  
10: gb\_sts.\*  
11: gb\_sy.\*  
12: gb\_un.\*  
13: gb\_vi.\*  
14: gb\_htg.\*  
15: gb\_pl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	22	100.0	22	6	CQ771180	CQ771180 Sequence
C 2	22	100.0	1116	6	CQ771160	CQ771160 Sequence
C 3	22	100.0	196203	8	AL355355	AL355355 Human DNA
4	22	100.0	208202	14	AC013446	AC013446 Homo sapi
5	19.4	88.2	4004	6	CQ715783	CQ715783 Sequence
6	19.4	88.2	4007	8	HSU70212	U70212 Human SIM1
C 7	19.4	88.2	20027	8	HSU319954	AL121948 Human DNA
8	19.4	88.2	162907	14	AC027499	AC027499 Homo sapi
9	19.4	88.2	176085	14	AL590309	AL590309 Homo sapi
C 10	19.4	88.2	205371	14	AC157157	AC157157 Bos tauru
C 11	18.8	85.5	148727	14	CR848833	CR848833 Danio rer
C 12	18.8	85.5	156893	5	AL954144	AL954144 Zebrafish
C 13	18.8	85.5	227929	14	AC094794	AC094794 Rattus no
C 14	18.4	83.6	50905	14	AC099853	AC099853 Mus muscu
15	18.4	83.6	64031	14	AC100990	AC100990 Mus muscu
C 16	18.4	83.6	80308	14	AC119858	AC119858 Mus muscu
C 17	18.4	83.6	168370	9	AC147104	AC147104 Mus muscu
C 18	18.4	83.6	192120	14	AC163486	AC163486 Mus muscu

19	18.4	83.6	203304	9	AC126258	AC126258 Mus muscu
C 20	18.4	83.6	208992	9	AC153881	AC153881 Mus muscu
21	18.4	83.6	212483	9	AC159476	AC159476 Mus muscu
22	18.4	83.6	214192	14	AC140497	AC140497 Mus muscu
23	18	81.8	196838	9	AC153867	AC153867 Mus muscu
24	18	81.8	217496	9	AC153630	AC153630 Mus muscu
C 25	18	81.8	217586	9	AC158678	AC158678 Mus muscu
C 26	17.8	80.9	4363	1	AB014686	AB014686 Streptoco
C 27	17.8	80.9	39915	8	AC010512	AC010512 Homo sapi
C 28	17.8	80.9	72147	14	AC021693	AC021693 Homo sapi
29	17.8	80.9	87619	14	AC016504	AC016504 Homo sapi
C 30	17.8	80.9	92290	14	AC022755	AC022755 Homo sapi
31	17.8	80.9	110000	14	AC114055_2	Continuation (3 of
C 32	17.8	80.9	110000	14	AC128511_2	Continuation (3 of
C 33	17.8	80.9	130912	14	AC119050	AL731573 Gallus ga
34	17.8	80.9	131465	8	AL731573	AL731573 Human DNA
C 35	17.8	80.9	140414	14	AC136548	AC136548 Rattus no
C 36	17.8	80.9	144242	8	AL139812	AL139812 Human DNA
C 37	17.8	80.9	145618	14	CR847900	CR847900 Danio rer
C 38	17.8	80.9	153400	5	BX649515	BX649515 Zebrafish
C 39	17.8	80.9	160230	9	BX072552	BX072552 Mouse DNA
40	17.8	80.9	166937	14	AC116927	AC116927 Rattus no
C 41	17.8	80.9	169031	14	BX936460	BX936460 Danio rer
42	17.8	80.9	170059	8	AC026780	AC026780 Homo sapi
C 43	17.8	80.9	171474	14	CR956418	CR956418 Sus scrof
C 44	17.8	80.9	176502	14	AC134706	AC134706 Rattus no
C 45	17.8	80.9	177787	8	AC133961	AC133961 Homo sapi

ALIGNMENTS

RESULT 1	CQ771180	Sequence 28 from Patent EP1388590.	22 bp	DNA	linear	PAT 04-MAR-2004
LOCUS	CQ771180	Sequence 28 from Patent EP1388590.	22 bp	DNA	linear	PAT 04-MAR-2004
DEFINITION	CQ771180	Sequence 28 from Patent EP1388590.	22 bp	DNA	linear	PAT 04-MAR-2004
ACCESSION	CQ771180	Sequence 28 from Patent EP1388590.	22 bp	DNA	linear	PAT 04-MAR-2004
VERSION	CQ771180.1	GI:45125313	22 bp	DNA	linear	PAT 04-MAR-2004
KEYWORDS	CQ771180.1	GI:45125313	22 bp	DNA	linear	PAT 04-MAR-2004
SOURCE	CQ771180.1	GI:45125313	22 bp	DNA	linear	PAT 04-MAR-2004
ORGANISM	CQ771180.1	GI:45125313	22 bp	DNA	linear	PAT 04-MAR-2004
REFERENCE	CQ771180.1	GI:45125313	22 bp	DNA	linear	PAT 04-MAR-2004
AUTHORS	CQ771180.1	GI:45125313	22 bp	DNA	linear	PAT 04-MAR-2004
TITLE	CQ771180.1	GI:45125313	22 bp	DNA	linear	PAT 04-MAR-2004
JOURNAL	CQ771180.1	GI:45125313	22 bp	DNA	linear	PAT 04-MAR-2004
FEATURES	CQ771180.1	GI:45125313	22 bp	DNA	linear	PAT 04-MAR-2004
source	CQ771180.1	GI:45125313	22 bp	DNA	linear	PAT 04-MAR-2004
1..22	CQ771180.1	GI:45125313	22 bp	DNA	linear	PAT 04-MAR-2004
/organism="synthetic construct"	CQ771180.1	GI:45125313	22 bp	DNA	linear	PAT 04-MAR-2004
/mol_type="unassigned DNA"	CQ771180.1	GI:45125313	22 bp	DNA	linear	PAT 04-MAR-2004
/db_xref="taxon:32630"	CQ771180.1	GI:45125313	22 bp	DNA	linear	PAT 04-MAR-2004

ORIGIN

Query Match	100.0%	Score 22;	DB 6;	Length 22;
Best Local Similarity	100.0%;	Pred. No. 4.6;	Mismatches 0;	Indels 0;
Matches 22;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	TCATGCTCACATTAACTGGA	22	
Db	1	TCATGCTCACATTAACTGGA	22	
RESULT 2	CQ771160/c	Sequence 8 from Patent EP1388590.	1116 bp	DNA
LOCUS	CQ771160/c	Sequence 8 from Patent EP1388590.	1116 bp	DNA
DEFINITION	CQ771160/c	Sequence 8 from Patent EP1388590.	1116 bp	DNA
ACCESSION	CQ771160/c	Sequence 8 from Patent EP1388590.	1116 bp	DNA
VERSION	CQ771160.1	GI:45125293	1116 bp	DNA
KEYWORDS	CQ771160.1	GI:45125293	1116 bp	DNA
SOURCE	CQ771160.1	GI:45125293	1116 bp	DNA
ORGANISM	CQ771160.1	GI:45125293	1116 bp	DNA
Homo sapiens (human)	CQ771160.1	GI:45125293	1116 bp	DNA
Homo sapiens	CQ771160.1	GI:45125293	1116 bp	DNA
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	CQ771160.1	GI:45125293	1116 bp	DNA

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

REFERENCE

1 Kouchi,Y., Masago,A. and Takahata,T.  
Gene assay method for predicting glaucoma onset risk  
Patent: EP 1388590-A 8 11-FEB-2004;  
Sysmex Corporation (JP)

FEATURES

source

Location/Qualifiers

1. .1116

/organism="Homo sapiens"

/mol\_type="unassigned DNA"

/db\_xref="taxon:9606"

ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 1116;

Best Local Similarity 100.0%; Pred. No. 3.3;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCATGCTCACACATTAACTGGA 22

|||||

Db 786 TCATGCTCACACATTAACTGGA 765

RESULT 3

AL355355/c

LOCUS

AL355355 196203 bp DNA linear PRI 18-MAY-2005  
Human DNA sequence from clone RP11-730A19 on chromosome 10 Contains  
the 5' end of a novel gene (DKFZP761F241) (FLJ20925 FLJ38473), a  
ribosomal protein L5 (RPL5) pseudogene, the OPTN gene for  
optineurin, a small nuclear ribonucleoprotein polypeptide G (SNRPG)  
pseudogene, a pseudogene similar to part of COX10 homolog,  
cytochrome c oxidase assembly protein heme A: farnesyltransferase  
(yeast) (COX10), a novel pseudogene (FLJ10648 KIAA1525), a  
ribosomal protein L36A (RPL36A) pseudogene, the 5' end of the  
MCM10 gene for MCM10 minichromosome maintenance deficient 10 (S.  
cerevisiae), a pseudogene similar to part of chromodomain protein Y  
chromosome 1 (CDY1) and two CpG islands, complete sequence.

ACCESSION

AL355355.25 GI:16972859

HTG; CDY1; chromodomain; COX10; CpG island; DKFZP761F241; FLJ10648;

FLJ20925; FLJ38473; KIAA1525; MCM10; optineurin; OPTN; RPL36A;

RPL5; SNRPG.

Source Homo sapiens (human)

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

1 (bases 1 to 196203)

REFERENCE

A Almeida,J.

Direct Submission

Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk

Clone requests: clonerequest@sanger.ac.uk

On Nov 16, 2001 this sequence version replaced gi:16214585.

The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:

Em.; EMBL; Sw.; SWISSPROT; Tr.; TREMBL; Wp.; WORKPEP; Information

on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence

was generated from part of bacterial clone contigs of human

chromosome 10, constructed by the Sanger Centre Chromosome 10

Mapping Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr10

RP11-730A19 is from the library RPCI-11.3 constructed by the group

of Pieter de Jong. For further details see

http://www.chori.org/bacpac/home.htm

VECTOR: pBac3.6

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: vegas@sanger.ac.uk

-----

This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one subclone; and the assembly was confirmed by restriction digest,  
except on the rare occasion of the clone being a YAC.

FEATURES

source

Location/Qualifiers

1. .196203

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="10"

/clone="RP11-730A19"

/clone\_lib="RPCI-11.3"

misc\_feature

1

gene

Join(Complement(123034..123180),

complement(122612..122837), complement(122053..122130),

complement(97396..97628), complement(73202..73469),

complement(21866..22040),

complement(AL353586.14:43768..45820))

/gene="RP11-347122.1"

/locus\_tag="RP11-347122.1-001"

Join(Complement(123034..123180),

complement(122612..122837), complement(122053..122130),

complement(97396..97628), complement(73202..73469),

complement(21866..22040),

complement(AL353586.14:43768..45820))

/gene="RP11-347122.1"

/locus\_tag="RP11-347122.1-001"

/product="novel protein"

/note="match: cDNAs: AK095792.1"

Join(Complement(24725..25225), complement(21866..22040),

complement(AL353586.14:43768..45820))

/gene="RP11-347122.1"

/locus\_tag="RP11-347122.1-002"

Join(Complement(24725..25225), complement(21866..22040),

complement(AL353586.14:43768..45820))

/gene="RP11-347122.1"

/locus\_tag="RP11-347122.1-002"

/product="novel protein"

/note="match: ESTs: AL514062.1 AL532821.1 AL552108.1

AL571318.1 AL571347.1 AW965893.1 BR299213.1 BF307636.1

BF332322.1 BF582825.1 BM503747.1 BM677538.1 BM983097.1

BQ044480.1 BQ331824.1 BQ447486.1 BQ720316.1 BQ722244.1

BQ723478.1 BQ894764.1 BQ897889.1 BU726400.1 BU901542.1

match: cDNAs: AB050523.1 AK009833.1 AL136562.1"

Join(Complement(24725..25098), complement(21866..22040),

complement(AL353586.14:45557..45820))

/gene="RP11-347122.1"

/locus\_tag="RP11-347122.1-002"

/standard\_name="OTTHUMP0000019122"

/note="match: proteins: Q9BQI4 Q9D6Y1"

/codon\_start=1

/product="novel protein"

/protein\_id="CAI16546.1"

/db\_xref="GI:55958689"

/translation="MLRQLLAALCLAGPPAPACQLPSEWRPLSEGCRAELAEIV

YARVALHPEAFGLYNIHFLPWYHAGGGLFYSAEVMCLDQAWGMLFVPGSRNLIT

GLGYFCSHSHTVVQDYVFFFLRMDBENTNLPLPHGVNFQDAIPFDQENRMFSSLPQF

SNCKVQSQQLATFSSDWEIQDSRLKSSVQKALFEEDHVKLQKVATLEKRNQLR

ERVKVKSIRQARKKGRHLELANOKLSKLAAGALPHINARGPVRPPYLRG"

Join(Complement(21866..22039),

complement(AL353586.14:45557..45820))

/gene="RP11-347122.1"

/locus\_tag="RP11-347122.1-001"

/standard\_name="OTTHUMP0000019121"

/codon\_start=1

/product="novel protein"

/protein\_id="CAI16545.1"

/db\_xref="GI:55958688"

/db\_xref="UniProt/TREMBL:Q5VVV9"

/translation="MDENTNLLPHGVNFQDAIFPDPTQENRRMFSSLFQPSNCSQGOQL  
ATFSDDWEQSDRLMSVQKALFEEDHVKKLOKQVATLEKRRNQLRERVKYKVRs  
LRQAQKHGHSLANQKLSEKLAAGALPHINARGPVRPPYLRG"  
59854. .59914  
/note="Sequence from overlapping clone RP11-141H1  
(AL596124)". Assembly confirmed by restriction digest."  
complement (81430. .82204)  
/locus\_tag="RP11-730A19.2-001"  
/pseudo  
complement (81430. .82204)  
/locus\_tag="RP11-730A19.2-001"  
/note="match: proteins: RAN05603 O22608 O44248 O65353  
O76190 P09895 F15125 F15126 P19949 P22451 P46777 P47962  
P49405 Q8MINS QBMTW3 Q90YW5 Q90YW6 Q95276 Q9AVH1 Q9H3F4  
Q9W5R8 Q9Y0H6"  
/pseudo  
/codon\_start=1  
/product="ribosomal protein L5 (RPL5) pseudogene"  
98305. .98335  
/note="Sequence from overlapping clone RP11-141H1  
(AL596124)". Assembly confirmed by restriction digest."  
121303  
/note="Clone right end: RP11-141H1"  
join(122977. .123486,131666. .131817,132640. .132816,  
133802. .134004,135981. .136163,137995. .139868,  
142416. .142568,145913. .146015,147523. .147638,  
148946. .149095,149474. .149567,151273. .151431,  
155595. .155725,157030. .157109,160273. .161819)  
/gene="OPTN"  
/locus\_tag="RP11-730A19.1-006"  
join(122977. .123486,131666. .131817,132640. .132816,  
133802. .134004,135981. .136163,137995. .139868,  
142416. .142568,145913. .146015,147523. .147638,  
148946. .149095,149474. .149567,151273. .151431,  
155595. .155725,157030. .157109,160273. .161819)  
/gene="OPTN"  
/locus\_tag="RP11-730A19.1-006"  
/product="optineurin"  
/note="match: ESTs: AA196381.1 AA442279.1 AI675996.1  
A1760653.1 AI937171.1 AL134550.1 AL562664.1 AL565831.1  
AL583163.1 AL583164.1 AL708106.1 AW156882.1 AW162062.1  
AW162138.1 AW297795.1 AW630059.1 BE875862.1 BE884512.1  
BF195338.1 BG026508.1 BG031732.1 BGI19942.1 BG222549.1  
BG427366.1 BG530095.1 BG678027.1 B1560200.1 B1770911.1  
BM079706.1 BM809924.1 BM992478.1 BQ719953.1 BU623376.1  
BU28287.1 BU733648.1 CA416446.1 N23490.1  
match: cDNAs: AK055403.1"  
join(123110. .123486,131666. .131817,132640. .132816,  
133802. .134004,135981. .136163,137995. .139868,  
142434. .142568,145913. .146015,147523. .147638,  
148946. .149095,149474. .149567,151273. .151431,  
155595. .155725,157030. .157109,160273. .161819)  
/gene="OPTN"  
/locus\_tag="RP11-730A19.1-009"  
join(123110. .123486,131666. .131817,132640. .132816,  
133802. .134004,135981. .136163,137995. .139868,  
142434. .142568,145913. .146015,147523. .147638,  
148946. .149095,149474. .149567,151273. .151431,  
155595. .155725,157030. .157109,160273. .161819)  
/gene="OPTN"  
/product="optineurin"  
/note="match: cDNAs: AB063036.1"  
join(123687. .123830,132640. .132816,133001. .133190)  
/gene="OPTN"  
/locus\_tag="RP11-730A19.1-010"  
join(123687. .123830,132640. .132816,133001. .133190)  
/gene="OPTN"  
/locus\_tag="RP11-730A19.1-010"  
/product="optineurin"  
/note="match: ESTs: AA182845.1 AA187695.1"  
join(123691. .123830,132640. .132816,135981. .136163,  
139795. .139868,142416. .142508)

/gene="OPTN"  
/locus\_tag="RP11-730A19.1-002"  
join(123691. .123830,132640. .132816,135981. .136163,  
139795. .139868,142416. .142508)  
/gene="OPTN"  
/locus\_tag="RP11-730A19.1-002"  
Query Match 100.0%; Score 22; DB 8; Length 196203;  
Best Local Similarity 100.0%; Pred. NO. 2.1;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TCATGCTCACACATTAACTGGA 22  
Db 147808 TCATGCTCACACATTAACTGGA 147787  
RESULT 4  
AC013446  
LOCUS  
DEFINITION Homo sapiens chromosome 10 clone RP11-513P21, WORKING DRAFT  
SEQUENCE, 23 unordered pieces.  
ACCESSION AC013446  
VERSION AC013446.3 GI:7923397  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo  
REFERENCE  
AUTHORS 1 (bases 1 to 208202)  
TITLE Waterston,R.H.  
JOURNAL The sequence of Homo sapiens clone  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 208202)  
AUTHORS Waterston,R.H.  
JOURNAL Direct Submission  
TITLE Submitted (11-NOV-1999) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
COMMENT  
On May 18, 2000 this sequence version replaced gi:6850545.  
----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc/index.shtml>  
----- Project Information -----  
Center project name: H NH0513P21  
----- Summary Statistics -----  
Sequencing vector: MI3; 84%  
Sequencing vector: plasmid; 16%  
Chemistry: Dye-primer ET; 84% of reads  
Chemistry: Dye-terminator Big Dye; 16% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 196339 bases at least Q40  
Consensus quality: 199886 bases at least Q30  
Consensus quality: 202038 bases at least Q20  
Insert size: 215000; agarose-fp  
Insert size: 206002; sum-of-contigs  
Quality coverage: 4.38 in Q20 bases; agarose-fp  
Quality coverage: 4.26 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 23 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 1699: contig of 1699 bp in length  
\* 1700 1799: gap of unknown length  
\* 1800 3240: contig of 1441 bp in length  
\* 3241 3240: gap of unknown length





Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCATGCTCACATTAAGTGA 22  
|||||  
Db 58458 TCATGCTCACATTAAGTGA 58479

RESULT 5  
LOCUS CQ715783  
DEFINITION Sequence 1717 from Patent WO02068579.  
ACCESSION CQ715783  
VERSION CQ715783.1 GI:42276640

SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

REFERENCE Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.  
AUTHORS Kits, such as nucleic acid arrays, comprising a majority of  
TITLE humanexons or transcripts, for detecting expression and other uses  
thereof

JOURNAL Patent: WO 02068579-A 1717 06-SEP-2002;  
PE Corporation (NY) (US)  
FEATURES Location/Qualifiers  
source 1..4004  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Query Match 88.2%; Score 19.4; DB 6; Length 4004;  
Best Local Similarity 95.2%; Pred. No. 64;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 CATGCTCACATTAAGTGA 22  
|||||  
Db 2340 CATGCTCACATTAAGTGA 2360

RESULT 6  
LOCUS HSU70212  
DEFINITION Human SIM1 mRNA, complete cds.  
ACCESSION U70212  
VERSION U70212.1 GI:2245351

SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

REFERENCE Chraat, R., Scott, H.S., Chen, H., Kudoh, J., Rossier, C., Minoshima, S.,  
AUTHORS Wang, Y., Shimizu, N. and Antonarakis, S.E.  
TITLE Cloning of two human homologs of the Drosophila single-minded gene  
SIM1 on chromosome 6q and SIM2 on 21q within the Down syndrome  
chromosomal region

JOURNAL Genome Res. 7 (6), 615-624 (1997)  
PUBMED 919934  
REFERENCE 2 (bases 1 to 4007)  
AUTHORS Chraat, R., Rossier, C. and Antonarakis, S.E.  
TITLE Direct Submission

JOURNAL Submitted (10-SEP-1996) Medical Genetics, Geneva University Medical  
School, 1 rue Michel Servet, Geneva 1211, Switzerland

FEATURES Location/Qualifiers  
source 1..4007  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="6"

/map="6q"  
1..4007  
/gene="SIM1"  
217..2517  
/gene="SIM1"  
/note="Similar to the single-minded protein in Drosophila  
melanogaster, SwissProt Accession Number P05709"  
/codon\_start=1  
/product="hSIM1"  
/protein\_id="AAB62395.1"  
/db\_xref="GI:2245352"

/translation="MKKSKNAARTREKENSEFYEELAKLLPLASAITSQVDKASIIIR  
LTTSLKMRVFPFEGELGAWSSRTSPLDNVGREGSLQLTQDFIFVAPDQKIM  
YISFASVHLGSLQVELTGSYIEIHPADHDEMTAVLTAHQYHSHFVQYIEIRSF  
FLRMKCVLAKRNAGLTCGYKVIHCGSLDKIQLKQYSLDMSFPDGCYQNVGLVAGSLP  
PSAVTEIKLHNSMFMFRASLDMLKIFLDSRVAELTGYRPODLIETLVHHVHCDTFH  
LRCAHLLLVKGVQVTKYRPLAKHGMVQSVATIVHNSPSPHICIVSNVYLTD  
TEYKGLQLSLDQISASKAPSYTSSSTPTMTDNKGAKSRLLSSKSKRTSPYQVSG  
FHTKSESDHDSQWGSPLTDTASPOLDDPADRPQSDHSDASCAIRQFSDRSLCYGFA  
LDHSLRVEERHFHTQACGRCAGRYFLGTPOAGREPWMSRAALPLTKAPESREA  
YENSMPIASVHRHGRGHWDSDSVSSPDGSGASGSDRYRTEQYQPPSPHPSKTEI  
LIRATQOMIKBEENRLQRLKAPSDOLASINGAKHSLCFANYQPPPTGVECHGSL  
LITRATQCDHIQOREGKWLSPHENDYDNPALSRISSPNSDRISKSLILAKDYLHSDI  
SPHQTAGHPHTVSPNCFSGSHRQYFDKHYATLTGYALEHLYDSETIRNYSGLCNGSHFD  
VTSHURMQDPDPAQGHKGYSVIITNGS"

ORIGIN  
Query Match 88.2%; Score 19.4; DB 8; Length 4007;  
Best Local Similarity 95.2%; Pred. No. 64;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 CATGCTCACATTAAGTGA 22  
|||||  
Db 2341 CATGCTCACATTAAGTGA 2361

RESULT 7  
LOCUS HSDJ399E4  
DEFINITION Human DNA sequence from clone RP3-399E4 on chromosome 6 Contains  
the 3' end of the SIM1 gene for single-minded (Drosophila) homolog  
1, ESTs, STSs and GSSs, complete sequence.

ACCESSION AL121948  
VERSION AL121948.8 GI:7981300  
KEYWORDS HTG; SIM1.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

REFERENCE 1 (bases 1 to 20027)  
AUTHORS Parker, A.  
TITLE Direct Submission  
JOURNAL Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk

COMMENT Clone requesters: clonerequest@sanger.ac.uk  
On May 22, 2000 this sequence version replaced gi:7106641.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em., EMBL; Swi., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information  
on the WORMPEP database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human

chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping  
Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr6

Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: vegas@sanger.ac.uk  
-----  
This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

RP3-399E4 is from the library RP3-3 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>  
VECTOR: pCYPAC2.

## FEATURES

```

source
    Location/Qualifiers
    1..20027
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
        /chromosomes="6"
        /clone="RP3-399E4"
        /clone_lib="RP3-3"
misc_feature
    1
        /note="Clone left end: RP3-399E4"
        join(complement(286062..1:20381)..63266),
        complement(4527..13407))
        /genes="SIM1"
        /locus tag="RP3-399E4.1-001"
        join(complement(286062..1:62885)..63266),
        complement(286062..1:53353)..53435),
        complement(286062..1:49854)..49947),
        complement(286062..1:49182)..49286),
        complement(286062..1:48954)..49039),
        complement(286062..1:48070)..48269),
        complement(286062..1:47737)..47843),
        complement(286062..1:46859)..47006),
        complement(286062..1:20381)..20549),
        complement(13005..13407),complement(4527..10609))
        /genes="SIM1"
        /locus tag="RP3-399E4.1-001"
        /note="match: ESTs: AA102407 AI011075 AI030699 AT241078
        A1573411 A1671360 AL036122 AL036129 AL598103 AW879919
        BE333766 BE663706 BE685233 BF284346 BF476109 BF672760
        BF673040 BF758122 BG428008 BG798852 BM506747 BM507096
        match: cDNAs: AF283298 AF363019 AY028626 G340058 HS001858
        HSM00312 HSU70212 HSU80456 HSU80457 MM40576
        MM42554 MMD135 MMD209 MMMSIMA"
        /genes="SIM1"
        /locus tag="RP3-399E4.1-001"
        complement(4547..4552)
        /genes="SIM1"
polyA_site
    1
        /locus tag="RP3-399E4.1-001"
        complement(4547..4552)
        /genes="SIM1"
polyA_signal
    1
        /locus tag="RP3-399E4.1-001"
        /genes="SIM1"
CDS
    1
        join(complement(286062..1:62885)..63059),
        complement(286062..1:53353)..53435),
        complement(286062..1:49854)..49947),
        complement(286062..1:49182)..49286),
        complement(286062..1:48954)..49039),
        complement(286062..1:48070)..48269),
        complement(286062..1:47737)..47843),
        complement(286062..1:46859)..47006),
        complement(286062..1:20381)..20549),
        complement(13005..13407),complement(9879..10609))
        /genes="SIM1"
        /locus tag="RP3-399E4.1-001"
        /standard_name="OTTHUMP00000016902"
        /note="match: proteins: Q35391 Q70284 P81133 Q14190 Q61045
        Q61079 Q8UV3 Q98SU5 Q9DDU6 Q9NTR8"
        /codon_start=1
        /protein_id="CA120423.1"
        /db_xref="GI:56203444"
        /db_xref="GOA:Q5TDP7"
        /db_xref="InterPro:IPR000014"
        /db_xref="InterPro:IPR001067"
        /db_xref="InterPro:IPR001092"
        /db_xref="InterPro:IPR001610"
        /db_xref="InterPro:IPR010578"

```

```

/db_xref="UniProt/TREMBL:Q5TDP7"
/translation="WKSKNAARTREKENSEFVELAKLLPLPSAITSOLDKASIIIR
LTTSLKRVVFPPEGLGEANGSSRTSPLDNVGRELGSLLQLTLDGFIIVVAPDGKIM
YISITASVHLGLSQVELTNGSIYIHPADHEMTAVLTATQPHSHFVQVEIERSF
FLRMKCVLAKNAGLTCGGYKVIHCISGLYKIRQYSLDMSPDFGQYQNVGLVAVGHSPL
PSAITEIKLSNMFMRASLDKLIFLDSRVAELTGYEPDQILKTLXHHVHGCDTFH
TECAHLLLVKGVQVTKYRFLAKHGWMVQVSATYVHNSRSSRPHCIVSNVYVLTQ
TEYKGLQSLDOIASKAPAFSVYSSSTPTMDNRKAKSLSSSKSKSRSTSPYPOYSG
FHTERSDHDHSQMGSPDITATSPQLLDPADRPQSHDASCAYQPSDRSLCYGFA
LDHSLVEERHFTACGECRAGRYFLGTPQAGREFPWSRAALPUTKASPSEREA
YENSMPHIAFVHRIHGRGHWDSESVSPDGSSESDRYRTQYQSSPHEPSKIET
LIRATQCMIKEENRLQIRKAPSDQLASINGAKKHSICFANYQQPPTGVCVCHSAL
LIRATPCQMIQOREGKMLSPHENDYDNSPTALSRISSPNSDRISKSLLILAKDYLHSDI
SPHOTAGDHPVSPNCFSGSHRQYFDKAYTITGYALEHLYDSEITRNYSLGCNGSHFD
VISHLRMQPDPQAQGHKGTSLVITNGS"
19928
/note="Clone_left_end: RP1-121G13"

```

misc\_feature

ORIGIN

Query Match 88.2%; Score 19.4; DB 8; Length 20027;  
 Best Local Similarity 95.2%; Pred. No. 55;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CATGCTCACATTAACTGGA 22

Db 10055 CATGTTACACATTAACTGGA 10035

RESULT 8

AC027499

LOCUS

AC027499 162907 bp DNA linear HTG 26-MAY-2000  
 Homo sapiens chromosome 6 clone RP11-699H20 map 6, WORKING DRAFT  
 SEQUENCE, 18 unordered pieces.

ACCESSION

AC027499.3

VERSION

HTG; HTGS\_PHASE1; HTGS\_DRAFT.

KEYWORDS

Homo sapiens (human)

SOURCE

Homo sapiens

ORGANISM

REFERENCE

1 (bases 1 to 162907)

AUTHORS

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE

Homo sapiens chromosome 6, clone RP11-699H20

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 162907)

AUTHORS

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bada,F.,

Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,

Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,

Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,

Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,

Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,

Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,

Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lechoczky,J.,

Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,

McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,

Meldrim,J., Meneus,L., Mihoval,T., Miranda,C., Mienna,V., Morrow,J.,

Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,

O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,

Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,

Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,

Stange-Rothmann,N., Stojanovic,N., Subramanian,A., Talamas,J.,

Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,

Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,

Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission

TITLE

JOURNAL

COMMENT

Submitted (30-MAR-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On May 25, 2000 this sequence version replaced gi:7656764.

All repeats were identified using RepeatMasker:

Smit, A.F.A. &amp; Green, P. (1996-1997)

```

http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8455
Center clone name: 699_H_20
----- Summary Statistics
Sequencing vector: ML3; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 152243 bases at least Q40
Consensus quality: 157677 bases at least Q30
Consensus quality: 159793 bases at least Q20
Insert size: 163000; agarose-fp
Insert size: 161207; sum-of-contigs
Quality coverage: 4.1 in Q20 bases; agarose-fp
Quality coverage: 4.2 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1039: contig of 1039 bp in length
* 1040 1139: gap of 100 bp
* 1140 3661: contig of 2522 bp in length
* 3662 3761: gap of 100 bp
* 3762 3792: contig of 3031 bp in length
* 3793 6892: gap of 100 bp
* 6893 10882: contig of 3990 bp in length
* 10883 14638: gap of 100 bp
* 14639 14737: contig of 3655 bp in length
* 14738 19171: contig of 4434 bp in length
* 19172 19271: gap of 100 bp
* 19272 23272: contig of 4001 bp in length
* 23273 23372: gap of 100 bp
* 23373 29778: contig of 6406 bp in length
* 29779 29878: gap of 100 bp
* 29879 33389: contig of 3511 bp in length
* 33390 33489: gap of 100 bp
* 33490 39967: contig of 6478 bp in length
* 39968 40067: gap of 100 bp
* 40068 47748: contig of 7681 bp in length
* 47749 47848: gap of 100 bp
* 47849 56708: contig of 8860 bp in length
* 56709 56808: gap of 100 bp
* 56809 65888: contig of 9080 bp in length
* 65889 65988: gap of 100 bp
* 65989 76847: contig of 10859 bp in length
* 76848 76947: gap of 100 bp
* 76948 92005: contig of 15058 bp in length
* 92006 92105: gap of 100 bp
* 92106 112123: contig of 20018 bp in length
* 112124 112223: gap of 100 bp
* 112224 131998: contig of 19775 bp in length
* 131999 132098: gap of 100 bp
* 132099 162907: contig of 30809 bp in length.
Location/Qualifiers
1. .162907
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="6"
/map="6"
/clone="RP11-699H20"
/clone_lib="RPC1-11 Human Male BAC"

```

FEATURES  
source

```

misc_feature 1. .1039
/note="assembly_fragment"
1040. .1139
/estimated_length=100
misc_feature 1140. .3661
/note="assembly_fragment"
3662. .3761
/estimated_length=100
misc_feature 3762. .6792
/note="assembly_fragment"
6793. .6892
/estimated_length=100
misc_feature 6893. .10882
/note="assembly_fragment"
10883. .10982
/estimated_length=100
misc_feature 10983. .14637
/note="assembly_fragment"
14638. .14737
/estimated_length=100
misc_feature 14738. .19171
/note="assembly_fragment"
19172. .19271
/estimated_length=100
misc_feature 19272. .23272
/note="assembly_fragment"
23273. .23372
/estimated_length=100
misc_feature 23373. .29778
/note="assembly_fragment"
29779. .29878
/estimated_length=100
misc_feature 29879. .33389
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
33390. .33489
/estimated_length=100
misc_feature 33490. .39967
/note="assembly_fragment"
39968. .40067
/estimated_length=100
misc_feature 40068. .47748
/note="assembly_fragment"
47749. .47848
/estimated_length=100
misc_feature 47849. .56708
/note="assembly_fragment"
56709. .56808
/estimated_length=100
misc_feature 56809. .65888
/note="assembly_fragment"
clone_end:17
vector_side:left"
65889. .65988
/estimated_length=100
misc_feature 65989. .76847
/note="assembly_fragment"
76848. .76947
/estimated_length=100
misc_feature 76948. .92005
/note="assembly_fragment"
92006. .92105
/estimated_length=100
misc_feature 92106. .112123
/note="assembly_fragment"
112124. .112223
/estimated_length=100
misc_feature 112224. .131998
/note="assembly_fragment"
131999. .132098
/estimated_length=100
misc_feature 132099. .162907

```

```
ORIGIN
/note="assembly_fragment"

Query Match      88.2%; Score 19.4; DB 14; Length 162907;
Best Local Similarity 95.2%; Pred. NO. 46;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 CATGCTCACACATTAACATCGA 22
Db      138532 CATGCTTACACATTAACATCGA 138552

RESULT 9
AL590309      176085 bp      DNA      linear      HTG 19-AUG-2001
LOCUS Homo sapiens chromosome 6 clone RP11-123D20, 39 unordered pieces.
DEFINITION AL590309
ACCESSION AL590309
VERSION AL590309.8 GI:15041899
KEYWORDS HTG; HTGS PHASE1; HTGS_CANCELLED.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
          Homnidae; Homo.

REFERENCE
1
AUTHORS Sims,S.
TITLE Direct Submission
JOURNAL Submitted (18-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
        CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
        requests: clonerequest@sanger.ac.uk
COMMENT On Jul 30, 2001 this sequence version replaced gi:15022265.
        ----- Genome Center
        Center: Sanger Centre
        Center code: SC
        Web site: http://www.sanger.ac.uk
        Contact: humquery@sanger.ac.uk
        ----- Project Information
        Center project name: BA123D20
        ----- Summary Statistics
        Assembly program: XGAP4; version 4.5
        Sequencing vector: plasmid; L08752; 100% of reads
        Chemistry: Dye-terminator Big Dye; 100% of reads
        Consensus quality: 163369 bases at least Q40
        Consensus quality: 168035 bases at least Q30
        Consensus quality: 170539 bases at least Q20
        Insert size: 172285; sum-of-contigs
        Insert size: 140408; agarose-fp
        Quality coverage: 4.10x in Q20 bases; sum-of-contigs Quality
        coverage: 7.26x in Q20 bases; agarose-fp
        -----
        * NOTE: This is a 'working draft' sequence. It currently
        * consists of 39 contigs. The true order of the pieces
        * is not known and their order in this sequence record is
        * arbitrary. Gaps between the contigs are represented as
        * runs of N, but the exact sizes of the gaps are unknown.
        * This record will be updated with the finished sequence
        * as soon as it is available and the accession number will
        * be preserved.
        *
        * 1 2735: contig of 2735 bp in length
        * 2736 2835: gap of 100 bp
        * 2836 11839: contig of 9004 bp in length
        * 11840 11939: gap of 100 bp
        * 11940 15478: contig of 3539 bp in length
        * 15479 15578: gap of 100 bp
        * 15579 18699: contig of 3121 bp in length
        * 18700 18799: gap of 100 bp
        * 18800 22632: contig of 3833 bp in length
        * 22633 22732: gap of 100 bp
        * 22733 25466: contig of 2734 bp in length
        * 25467 25566: gap of 100 bp
        * 25567 28259: contig of 2693 bp in length
        * 28260 28359: gap of 100 bp
        * 28360 35356: contig of 6997 bp in length

FEATURES
source
1..176085
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chrnosome="6"
/clone="RP11-123D20"
/clone_lib="RPC1-11.1"
1..2735
/note="assembly_fragment:04659
fragment_chain:1"

misc_feature
1..2735
/note="assembly_fragment:04659
fragment_chain:1"
```

misc\_feature 2836. .11839  
/note="assembly\_fragment:01821"  
fragment\_chain:1"  
misc\_feature 11940. .15478  
/note="assembly\_fragment:03816"  
fragment\_chain:1"  
misc\_feature 15579. .18699  
/note="assembly\_fragment:00904"  
fragment\_chain:2"  
misc\_feature 18800. .22632  
/note="assembly\_fragment:04044"  
fragment\_chain:2"  
misc\_feature 22733. .25466  
/note="assembly\_fragment:01880"  
fragment\_chain:3"  
misc\_feature 25567. .28259  
/note="assembly\_fragment:02106"  
fragment\_chain:3"  
misc\_feature 28360. .35356  
/note="assembly\_fragment:04673"  
fragment\_chain:4"  
misc\_feature 35457. .38329  
/note="assembly\_fragment:01162"  
fragment\_chain:4"  
misc\_feature 38430. .41049  
/note="assembly\_fragment:00050"  
misc\_feature 41150. .43431  
/note="assembly\_fragment:00203"  
misc\_feature 43532. .46388  
/note="assembly\_fragment:00889"  
misc\_feature 46489. .52558  
/note="assembly\_fragment:00978"  
misc\_feature 52659. .54842  
/note="assembly\_fragment:01150"  
misc\_feature 54943. .61793  
/note="assembly\_fragment:01253"  
misc\_feature 61894. .64910  
/note="assembly\_fragment:01273"  
misc\_feature 65011. .71804  
/note="assembly\_fragment:01370"  
misc\_feature 71905. .75626  
/note="assembly\_fragment:01449"  
misc\_feature 75727. .79483  
/note="assembly\_fragment:01450"  
misc\_feature 79584. .82432  
/note="assembly\_fragment:01654"  
misc\_feature 82533. .85497  
/note="assembly\_fragment:01840"  
misc\_feature 85598. .99488  
/note="assembly\_fragment:01891"  
misc\_feature 99589. .106961  
/note="assembly\_fragment:02460"  
misc\_feature 107062. .109126  
/note="assembly\_fragment:02794"  
misc\_feature 109227. .120734  
/note="assembly\_fragment:02802"  
misc\_feature 120835. .123057  
/note="assembly\_fragment:02934"  
misc\_feature 123158. .127772  
/note="assembly\_fragment:02949"  
misc\_feature 127873. .133773  
/note="assembly\_fragment:02976"  
misc\_feature 133874. .136606  
/note="assembly\_fragment:03101"  
misc\_feature 136707. .140974  
/note="assembly\_fragment:03201"  
misc\_feature 141075. .150433  
/note="assembly\_fragment:03389"

Query Match 88.2%; Score 19.4; DB 14; Length 176085;  
Best Local Similarity 95.2%; Pred. No. 46;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CATGCTCACATTAACTGGA 22  
|||||  
Db 167123 CATGCTTACATTAACTGGA 167143

RESULT 10  
AC157157/c  
LOCUS  
DEFINITION  
AC157157  
VERSION  
HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
KEYWORDS  
SOURCE  
Bos taurus (cow)  
ORGANISM  
Bos taurus

AC157157 205371 bp DNA linear HTG 01-JUL-2005  
Bos taurus clone CH240-64E14, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 15  
unordered pieces.

AC157157  
HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
Bos taurus (cow)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;  
Pecora; Bovidae; Bovinae; Bos.

1 (bases 1 to 205371)

Muzny, D. Marie., Metzker, M. Lee., Abramzon, S., Adams, C., Alder, J.,  
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,  
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
Bryant, N., Buhay, C., Burch, P., Burrell, K., Caesar, H., Calderon, E.,  
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,  
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,  
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
Georgiev, B., Geer, K., Gill, R., Grady, M., Guerrero, W., Guevara, M.,  
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,  
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,  
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,  
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
Lorensu, H., Loulsee, H., Lozada, R. J., Lu, X., Ma, J.,  
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,  
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
Mawney, S., McLeod, M. P., McNeill, T. Z., Meenen, S.,  
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nait, L.,  
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
Nwaokelemeh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,  
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,  
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, B. L.,  
Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,  
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,  
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,  
Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C. D., Smajls, D.,  
Snead, A., Sodergren, E., Song, X. Z., Sorelle, R., Sosa, J.,  
Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,  
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K.,  
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,  
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K.,  
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,  
Weinstock, G. and Gibbs, R. A.

Direct Submission  
Unpublished  
2 (bases 1 to 205371)  
Worley, K. C.  
Direct Submission

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE

## JOURNAL

Submitted (09-FEB-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 205371)  
Cow Genome Sequencing Consortium.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Direct Submission  
Submitted (01-JUL-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

## COMMENT

On Jun 29, 2005 this sequence version replaced gi:58801683. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information

Center project name: FPTH  
Center clone name: CH240-64E14  
----- Summary Statistics

Assembly program: Atlas 3.0;  
Consensus quality: 19569 bases at least Q40  
Consensus quality: 201314 bases at least Q30  
Consensus quality: 202773 bases at least Q20  
Estimated insert size: 204868; sum-of-contigs estimation  
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 15 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 11503: contig of 11503 bp in length  
\* 11504 11533: gap of 50 bp  
\* 11554 23036: contig of 11483 bp in length  
\* 23037 23086: gap of 50 bp  
\* 23087 27569: contig of 4483 bp in length  
\* 27570 27619: gap of 50 bp  
\* 27620 81122: contig of 53503 bp in length  
\* 81123 81172: gap of 50 bp  
\* 81173 93381: contig of 12209 bp in length  
\* 93382 93431: gap of 50 bp  
\* 93432 125400: contig of 31969 bp in length  
\* 125401 125450: gap of 50 bp  
\* 125451 166634: contig of 41184 bp in length  
\* 166635 166684: gap of 50 bp  
\* 166685 183593: contig of 16909 bp in length  
\* 183594 183643: gap of 50 bp  
\* 183644 195494: contig of 11851 bp in length  
\* 195495 195499: gap of unknown length  
\* 195500 197385: contig of 1791 bp in length  
\* 197386 197486: gap of unknown length  
\* 197487 198559: contig of 1073 bp in length  
\* 198560 200913: gap of unknown length  
\* 200914 201013: contig of 2255 bp in length  
\* 201014 202624: contig of 1611 bp in length

\* 202625 202724: gap of unknown length  
\* 202725 204214: contig of 1490 bp in length  
\* 204215 204315: gap of unknown length  
\* 204316 205371: contig of 1057 bp in length.

FEATURES  
source

1. .205371  
/organism="Bos taurus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9913"  
/clone="CH240-64E14"  
11504. .11553  
/estimated\_length=50  
23037. .23086  
/estimated\_length=50  
27570. .27619  
/estimated\_length=50  
81123. .81172  
/estimated\_length=50  
93382. .93431  
/estimated\_length=50  
125401. .125450  
/estimated\_length=50  
166635. .166684  
/estimated\_length=50  
183594. .183643  
/estimated\_length=50  
195495. .195499  
/estimated\_length=unknown  
197386. .197485  
/estimated\_length=unknown  
198559. .198658  
/estimated\_length=unknown  
200914. .201013  
/estimated\_length=unknown  
202625. .202724  
/estimated\_length=unknown  
204215. .204314  
/estimated\_length=unknown

## ORIGIN

Query Match 88.2%; Score 19.4; DB 14; Length 205371;  
Best Local Similarity 95.2%; Pred. No. 45;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTAACTGGA 22

Db 131438 CATGCTTACACATTAACTGGA 131418

RESULT 11  
CR848833/c

LOCUS 148727 bp DNA linear HTG 19-OCT-2004  
DEFINITION Danio rerio clone DKEY-126N12, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 21  
unordered pieces.

ACCESSION CR848833

VERSION CR848833.1 GI:54305964

KEYWORDS HTG; HTGS\_PHASE1.

SOURCE Danio rerio (zebrafish)

## ORGANISM

Danio rerio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
Cypriniformes; Cyprinidae; Danio.  
1 (bases 1 to 148727)

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

SimS, S.  
Direct Submission  
Submitted (18-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
fish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: [zfish-help@sanger.ac.uk](mailto:zfish-help@sanger.ac.uk)

```

----- Project Information
Center project name: zk126N12
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 140027 bases at least Q40
Consensus quality: 142485 bases at least Q30
Consensus quality: 144167 bases at least Q20
Insert size: 146727; sum-of-contigs
Quality coverage: 3.67x in Q20 bases; sum-of-contigs Quality
coverage: 3.59x in Q20 bases; agarose-gp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 2031: contig of 2031 bp in length
*
2032 2131: gap of 100 bp
*
2132 22619: contig of 20488 bp in length
*
22620 22719: gap of 100 bp
*
22720 25548: contig of 2829 bp in length
*
25549 25648: gap of 100 bp
*
25649 43931: contig of 18283 bp in length
*
43932 44032: gap of 100 bp
*
44032 64647: contig of 20616 bp in length
*
64648 64747: gap of 100 bp
*
64748 68231: contig of 3484 bp in length
*
68232 72399: contig of 4068 bp in length
*
72400 72499: gap of 100 bp
*
72500 75499: contig of 2999 bp in length
*
75500 75599: gap of 100 bp
*
75600 79180: contig of 3582 bp in length
*
79181 83340: contig of 4060 bp in length
*
83341 86629: contig of 3189 bp in length
*
86630 86729: gap of 100 bp
*
86730 90705: contig of 3976 bp in length
*
90706 90806: gap of 100 bp
*
90807 98871: contig of 8066 bp in length
*
98872 98971: gap of 100 bp
*
98972 103714: contig of 4743 bp in length
*
103715 103815: gap of 100 bp
*
103816 106612: contig of 2798 bp in length
*
106613 106713: gap of 100 bp
*
106714 109303: contig of 2591 bp in length
*
109304 109404: gap of 100 bp
*
109405 113779: contig of 4376 bp in length
*
113780 113880: gap of 100 bp
*
113881 116333: contig of 2454 bp in length
*
116334 116434: gap of 100 bp
*
116435 119180: contig of 2747 bp in length
*
119181 119281: gap of 100 bp
*
119282 126979: contig of 7699 bp in length
*
126980 127079: gap of 100 bp
*
127080 148727: contig of 21648 bp in length.
-----
FEATURES
source
1..148727
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEY-126N12"
/clone_lib="DanioKey"
1..2031
/notes="assembly_fragment:00112
fragment_chain:1"
2132..22619
misc_feature
10954144
AL954144
DEFINITION Zebrafish DNA sequence from clone CH211-206L15, complete sequence.
ACCESSION AL954144
VERSION AL954144.6 GI:28564318
KEYWORDS HTG.

```

```

/note="assembly_fragment:00958
fragment_chain:1"
22720..25548
/note="assembly_fragment:00125
fragment_chain:1"
25649..43931
/note="assembly_fragment:00631
fragment_chain:1"
44032..64647
/note="assembly_fragment:00500
fragment_chain:1"
64748..68231
/note="assembly_fragment:00323
fragment_chain:1"
68332..72399
/note="assembly_fragment:00181
fragment_chain:1"
72500..75498
/note="assembly_fragment:00237
fragment_chain:1"
75599..79180
/note="assembly_fragment:00198
fragment_chain:1"
79281..83340
/note="assembly_fragment:00300
fragment_chain:1"
83441..86629
/note="assembly_fragment:00090
fragment_chain:2"
86730..90705
/note="assembly_fragment:00257
fragment_chain:2"
90806..98871
/note="assembly_fragment:00388
fragment_chain:2"
98972..103714
/note="assembly_fragment:00346
fragment_chain:2"
103815..106612
/note="assembly_fragment:00151
fragment_chain:2"
106713..109303
/note="assembly_fragment:00165
fragment_chain:2"
109404..113779
/note="assembly_fragment:00278
fragment_chain:2"
113880..116333
/note="assembly_fragment:00052.0"
116434..119180
/note="assembly_fragment:00138"
119281..126979
/note="assembly_fragment:00440.0"
127080..148727
/note="assembly_fragment:00785"

ORIGIN
Query Match 85.5%; Score 18.8; DB 14; Length 148727;
Best Local Similarity 90.9%; Pred.No. 94;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCATGCTCACATTAACCTGGA 22
|||||
DB 84425 TCATCTCACATTAACCTGGA 84404

RESULT 12
AL954144/c
LOCUS AL954144 156893 bp DNA linear VRT 25-FEB-2003
DEFINITION Zebrafish DNA sequence from clone CH211-206L15, complete sequence.
ACCESSION AL954144
VERSION AL954144.6 GI:28564318
KEYWORDS HTG.

```



SOURCE	Danio rerio (zebrafish)	REFERENCE	
ORGANISM	Danio rerio	AUTHORS	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.		
REFERENCE	1 (bases 1 to 156893)		
AUTHORS	Phillimore,B.		
TITLE	Direct Submission		
JOURNAL	Submitted (25-FEB-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Feb 25, 2003 this sequence version replaced gi:28467501.		
COMMENT	----- Genome Center Center: Wellcome Trust Sanger Institute Center code: SC Web site: http://www.sanger.ac.uk Contact: zfish-help@sanger.ac.uk ----- During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/projects/C_elegans/wormpep Repeat names beginning 'Dr' were identified by The Recon repeat discovery system (Zhifeng Bao and Sean Eddy, submitted), and those beginning 'drr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see http://www/Projects/D_rerio/fishmask.shtml CH211-206L15 is from a CHORI-211 BAC library VECTOR: pTARBAC2.1		
FEATURES	Location/Qualifiers 1..156893 /organism="Danio rerio" /mol_type="genomic DNA" /db_xref="taxon:7955" /clone="CH211-206L15" /clone_lib="CHORI-211"		
ORIGIN			
Query Match	85.5%; Score 18.8; DB 5; Length 156893;		
Best Local Similarity	90.9%; Pred. No. 94;		
Matches	20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;		
Qy	1 TCATGCTCACATTAACCTGA 22 		
Db	118548 TCAATCTCACATTAACCTGA 118527		
RESULT 13			
AC094794/c			
LOCUS	Rattus norvegicus clone CH230-4N15, *** SEQUENCING IN PROGRESS ***.		
DEFINITION	AC094794		
ACCESSION	AC094794		
VERSION	AC094794.4 GI:30466624		
KEYWORDS	HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.		
SOURCE	Rattus norvegicus (Norway rat)		
ORGANISM	Rattus norvegicus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus.		

REFERENCE	1 (bases 1 to 227929)		
AUTHORS	Muzny,D., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J.J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,L., Cockrell,K., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.C., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mahoney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokeme,O., Okwomu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Ugmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wlaczky,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.		
	Direct Submission		
	Unpublished		
	2 (bases 1 to 227929)		
	Worley,K.C.		
	Direct Submission		
	Submitted (15-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA		
	3 (bases 1 to 227929)		
	Rat Genome Sequencing Consortium.		
	Direct Submission		
	Submitted (09-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA		
	On May 9, 2003 this sequence version replaced gi:22771587.		
	The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence		
COMMENT			



contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information

Center project name: GBJY

Center clone name: CH230-4N15

----- Summary Statistics

Assembly program: Atlas;

Consensus quality: 200503 bases at least Q40

Consensus quality: 205667 bases at least Q30

Consensus quality: 209426 bases at least Q20

Estimated insert size: 217540; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

-----

- \* NOTE: Estimated insert size may differ from sequence length (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).
- \* NOTE: This is a 'working draft' sequence. It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submittor.
- \* This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.
- \* 1 227929: contig of 227929 bp in length.

-----

Location/Qualifiers

1. .227929

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clone="CH230-4N15"

1. .3298

/notes="wgs end extension"

clone\_end:Sp6

5495. .6182

/notes="clone boundary"

clone\_end:Sp6

site:EcoRI

end\_sequence: BH309552"

-----

-----

Query Match 85.5%; Score 18.8; DB 14; Length 227929;

Best Local Similarity 90.9%; Pred. No. 91;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

-----

#### FEATURES

##### source

1. .227929

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clone="CH230-4N15"

##### misc\_feature

1. .3298

/notes="wgs end extension"

clone\_end:Sp6

5495. .6182

/notes="clone boundary"

clone\_end:Sp6

site:EcoRI

end\_sequence: BH309552"

-----

#### ORIGIN

Query Match 85.5%; Score 18.8; DB 14; Length 227929;

Best Local Similarity 90.9%; Pred. No. 91;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

-----

QY 1 TCATGCTCACATTAACTGGA 22

|||||

Db 88566 TCATGCTCACATTAACTGCA 88545

-----

#### RESULT 14

##### AC099853

LOCUS AC099853 50905 bp DNA linear HTG 22-NOV-2001

DEFINITION Mus musculus clone RP23-3N16, LOW-PASS SEQUENCE SAMPLING.

ACCESSION AC099853

VERSION AC099853.1 GI:17047219

KEYWORDS HTG; HTGS PHASE0.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 50905)

REFERENCE Birren, B., Linton, L., Nusbaum, C. and Lander, E.

TITLE Mus musculus, clone RP23-3N16

JOURNAL Unpublished

#### REFERENCE

##### AUTHORS

2 (bases 1 to 50905)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choe, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence.submissions@genome.wi.mit.edu](mailto:sequence.submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L3349

Center clone name: 3\_N16

-----

\* NOTE: This record contains 63 individual

\* sequencing reads that have not been assembled into

\* contigs. Runs of N are used to separate the reads

\* and the order in which they appear is completely

\* arbitrary. Low-pass sequence sampling is useful for

\* identifying clones that may be gene-rich and allows

\* overlap relationships among clones to be deduced.

\* However, it should not be assumed that this clone

\* will be sequenced to completion. In the event that

\* the record is updated, the accession number will

\* be preserved.

\* 1 693: contig of 693 bp in length

\* 694 793: gap of 100 bp

\* 794 1493: contig of 700 bp in length

\* 1494 1593: gap of 100 bp

\* 1594 2277: contig of 684 bp in length

\* 2278 2377: gap of 100 bp

\* 2378 3040: contig of 663 bp in length

\* 3041 3140: gap of 100 bp

\* 3141 3852: contig of 712 bp in length

\* 3853 3953: gap of 100 bp

\* 3953 4696: contig of 744 bp in length

\* 4697 4796: gap of 100 bp

\* 4797 5515: contig of 719 bp in length

\* 5516 5616: gap of 100 bp

\* 5616 6331: contig of 716 bp in length

\* 6332 6431: gap of 100 bp

\* 6432 7137: contig of 706 bp in length

\* 7138 7237: gap of 100 bp

\* 7238 7986: contig of 749 bp in length

\* 7987 8086: gap of 100 bp

\* 8087 8784: contig of 698 bp in length

\* 8785 8884: gap of 100 bp

\* 8885 9557: contig of 673 bp in length

\* 9558 9657: gap of 100 bp

\* 9658 9657: gap of 100 bp

\* 9658 10396: contig of 739 bp in length  
\* 10397 10496: gap of 100 bp  
\* 10497 11164: contig of 668 bp in length  
\* 11165 11264: gap of 100 bp  
\* 11265 11975: contig of 711 bp in length  
\* 11976 12075: gap of 100 bp  
\* 12076 12801: contig of 726 bp in length  
\* 12802 12901: gap of 100 bp  
\* 12902 13614: contig of 713 bp in length  
\* 13615 13714: gap of 100 bp  
\* 13715 14426: contig of 712 bp in length  
\* 14427 14526: gap of 100 bp  
\* 14527 15219: contig of 693 bp in length  
\* 15220 15319: gap of 100 bp  
\* 15320 15987: contig of 668 bp in length  
\* 15988 16087: gap of 100 bp  
\* 16088 16822: contig of 735 bp in length  
\* 16823 16922: gap of 100 bp  
\* 16923 17753: contig of 831 bp in length  
\* 17754 17853: gap of 100 bp  
\* 17854 18549: contig of 696 bp in length  
\* 18550 18649: gap of 100 bp  
\* 18650 19333: contig of 684 bp in length  
\* 19334 19433: gap of 100 bp  
\* 19434 20144: contig of 711 bp in length  
\* 20145 20244: gap of 100 bp  
\* 20245 20978: contig of 734 bp in length  
\* 20979 21078: gap of 100 bp  
\* 21079 21771: contig of 693 bp in length  
\* 21772 21871: gap of 100 bp  
\* 21872 22553: contig of 682 bp in length  
\* 22554 22653: gap of 100 bp  
\* 22654 23403: contig of 750 bp in length  
\* 23404 23503: gap of 100 bp  
\* 23504 24337: contig of 834 bp in length  
\* 24338 24437: gap of 100 bp  
\* 24438 25134: contig of 697 bp in length  
\* 25135 25234: gap of 100 bp  
\* 25235 25946: contig of 712 bp in length  
\* 25947 26046: gap of 100 bp  
\* 26047 26726: contig of 680 bp in length  
\* 26727 26826: gap of 100 bp  
\* 26827 27555: contig of 729 bp in length  
\* 27556 27655: gap of 100 bp  
\* 27656 28351: contig of 696 bp in length  
\* 28352 28451: gap of 100 bp  
\* 28452 29147: contig of 696 bp in length  
\* 29148 29247: gap of 100 bp  
\* 29248 29938: contig of 691 bp in length  
\* 29939 30038: gap of 100 bp  
\* 30039 30768: contig of 730 bp in length  
\* 30769 30868: gap of 100 bp  
\* 30869 31576: contig of 708 bp in length  
\* 31577 31676: gap of 100 bp  
\* 31677 32349: contig of 673 bp in length  
\* 32350 32449: gap of 100 bp  
\* 32450 33123: contig of 676 bp in length  
\* 33124 33225: gap of 100 bp  
\* 33226 33941: contig of 716 bp in length  
\* 33942 34041: gap of 100 bp  
\* 34042 34761: contig of 720 bp in length  
\* 34762 34861: gap of 100 bp  
\* 34862 35587: contig of 726 bp in length  
\* 35588 35687: gap of 100 bp  
\* 35688 36383: contig of 696 bp in length  
\* 36384 37163: contig of 680 bp in length  
\* 37164 37263: gap of 100 bp  
\* 37264 37940: contig of 677 bp in length  
\* 37941 38040: gap of 100 bp  
\* 38041 38755: contig of 715 bp in length  
\* 38756 38855: gap of 100 bp  
\* 38856 39535: contig of 680 bp in length

\* 39536 39635: gap of 100 bp  
\* 39636 40342: contig of 707 bp in length  
\* 40343 40442: gap of 100 bp  
\* 40443 41149: contig of 707 bp in length  
\* 41150 41249: gap of 100 bp  
\* 41250 41948: contig of 699 bp in length  
\* 41949 42048: gap of 100 bp  
\* 42049 42744: contig of 696 bp in length  
\* 42745 42844: gap of 100 bp  
\* 42845 43502: contig of 658 bp in length  
\* 43503 43602: gap of 100 bp  
\* 43603 44422: contig of 820 bp in length  
\* 44423 44522: gap of 100 bp  
\* 44523 45244: contig of 722 bp in length  
\* 45245 45344: gap of 100 bp  
\* 45345 46072: contig of 728 bp in length  
\* 46073 46172: gap of 100 bp  
\* 46173 46860: contig of 688 bp in length  
\* 46861 46960: gap of 100 bp  
\* 46961 47658: contig of 698 bp in length  
\* 47659 47758: gap of 100 bp  
\* 47759 48460: contig of 702 bp in length  
\* 48461 48560: gap of 100 bp  
\* 48561 49284: contig of 724 bp in length  
\* 49285 49384: gap of 100 bp  
\* 49385 50095: contig of 711 bp in length  
\* 50096 50195: gap of 100 bp  
\* 50196 50905: contig of 710 bp in length.

FEATURES  
source

1..50905  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/clone="RP23-3N16"  
/clone\_lib="RPCI-23 Female Mouse BAC"  
694..793  
/estimated\_length=100  
1494..1593  
/estimated\_length=100  
2278..2377

Query Match 83.6%; Score 18.4; DB 14; Length 50905;  
Best Local Similarity 95.0%; Pred. No. 1.7e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCATGCTCACATTAACGTG 20  
|||||  
Db 37744 TCATGCTCACATTAACGTG 37763

RESULT 15

AC100990  
LOCUS Mus musculus clone RP23-77K8, LOW-PASS SEQUENCE SAMPLING.  
DEFINITION AC100990  
ACCESSION AC100990  
VERSION AC100990.1 GI:17059764  
KEYWORDS HTG; HTGS\_PHASE0.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 64091)  
REFERENCE Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
TITLE Mus musculus, clone RP23-77K8  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 64091)  
AUTHORS Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhalter,B.,  
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,  
Choepel,Y., Collangelo,M., Collins,S., Collymore,A., Cook,A.,  
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,  
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,

Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,  
Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,  
Lamarez, R., Landers, T., Lehotsky, J., Levine, R., Liu, G.,  
MacLean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C.,  
McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J.,  
Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,  
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,  
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,  
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R.,  
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,  
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.P.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence.submissions@genome.wi.mit.edu](mailto:sequence.submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L14691  
Center clone name: 77\_K\_8  
-----  
\* NOTE: This record contains 78 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1  
686 785: contig of 685 bp in length  
786 785: gap of 100 bp  
1526 1525: contig of 740 bp in length  
1526 1625: gap of 100 bp  
1626 2366: contig of 741 bp in length  
2367 2466: gap of 100 bp  
3163 3262: contig of 696 bp in length  
3263 3969: contig of 707 bp in length  
3970 4069: gap of 100 bp  
4070 4793: contig of 724 bp in length  
4794 4893: gap of 100 bp  
4894 5609: contig of 716 bp in length  
5610 5709: gap of 100 bp  
5710 6439: contig of 730 bp in length  
6440 6539: gap of 100 bp  
6540 7269: contig of 730 bp in length  
7270 7369: gap of 100 bp  
7370 8093: contig of 724 bp in length  
8094 8193: gap of 100 bp  
8194 8907: contig of 714 bp in length  
8908 9007: gap of 100 bp  
9008 9711: contig of 704 bp in length  
9712 9811: gap of 100 bp  
9812 10521: contig of 710 bp in length  
10522 11334: gap of 100 bp  
11335 11434: contig of 713 bp in length  
11435 12139: contig of 705 bp in length  
12140 12239: gap of 100 bp  
12240 12970: contig of 731 bp in length

12971 13070: gap of 100 bp  
13071 13797: contig of 727 bp in length  
13798 13897: gap of 100 bp  
13898 14624: contig of 727 bp in length  
14625 14724: gap of 100 bp  
14725 15462: contig of 738 bp in length  
15463 15562: gap of 100 bp  
15563 16263: contig of 701 bp in length  
16264 16363: gap of 100 bp  
16364 17101: contig of 738 bp in length  
17102 17201: gap of 100 bp  
17202 17941: contig of 740 bp in length  
17942 18041: gap of 100 bp  
18042 18750: contig of 709 bp in length  
18751 18850: gap of 100 bp  
18851 19559: contig of 709 bp in length  
19560 19659: gap of 100 bp  
19660 20380: contig of 721 bp in length  
20381 20480: gap of 100 bp  
20481 21211: contig of 731 bp in length  
21212 21311: gap of 100 bp  
21312 22050: contig of 739 bp in length  
22051 22350: gap of 100 bp  
22351 22892: contig of 742 bp in length  
22893 22992: gap of 100 bp  
22993 23695: contig of 703 bp in length  
23696 23795: gap of 100 bp  
23796 24545: contig of 750 bp in length  
24546 24646: gap of 100 bp  
24646 25382: contig of 736 bp in length  
25383 25481: gap of 100 bp  
25482 26189: contig of 708 bp in length  
26190 26289: gap of 100 bp  
26290 26997: contig of 708 bp in length  
26998 27097: gap of 100 bp  
27098 27807: contig of 710 bp in length  
27808 28633: contig of 726 bp in length  
28634 28733: gap of 100 bp  
28734 29454: contig of 721 bp in length  
29455 29554: gap of 100 bp  
29555 30290: contig of 736 bp in length  
30291 30390: gap of 100 bp  
30391 31119: contig of 729 bp in length  
31120 31219: gap of 100 bp  
31220 31955: contig of 736 bp in length  
31956 32055: gap of 100 bp  
32056 32797: contig of 741 bp in length  
32797 32896: gap of 100 bp  
32897 33627: contig of 731 bp in length  
33628 33727: gap of 100 bp  
33728 34466: contig of 738 bp in length  
34466 34565: gap of 100 bp  
34566 35292: contig of 727 bp in length  
35293 35392: gap of 100 bp  
35393 36105: contig of 713 bp in length  
36106 36205: gap of 100 bp  
36206 36910: contig of 705 bp in length  
36911 37010: gap of 100 bp  
37011 37723: contig of 713 bp in length  
37724 37823: gap of 100 bp  
37824 38541: contig of 718 bp in length  
38542 38641: gap of 100 bp  
38642 39370: contig of 729 bp in length  
39371 39470: gap of 100 bp  
39471 40185: contig of 715 bp in length  
40186 40285: gap of 100 bp  
40286 41023: contig of 738 bp in length  
41024 41123: gap of 100 bp  
41124 41824: contig of 701 bp in length  
41825 41924: gap of 100 bp  
41925 42668: contig of 744 bp in length  
42669 42768: gap of 100 bp

TITLE  
JOURNAL  
COMMENT

```

* 42769 43491: contig of 723 bp in length
* 43492 43591: gap of 100 bp
* 43592 44303: contig of 712 bp in length
* 44304 44403: gap of 100 bp
* 44404 45128: contig of 725 bp in length
* 45129 45228: gap of 100 bp
* 45229 45958: contig of 730 bp in length
* 45959 46058: gap of 100 bp
* 46060 46797: contig of 738 bp in length
* 46798 46896: gap of 100 bp
* 46897 47640: contig of 744 bp in length
* 47641 48473: gap of 100 bp
* 48474 48573: contig of 733 bp in length
* 48574 49276: contig of 703 bp in length
* 49277 49376: gap of 100 bp
* 49377 50090: contig of 714 bp in length
* 50091 50902: contig of 712 bp in length
* 50903 51002: gap of 100 bp
* 51003 51726: contig of 724 bp in length
* 51727 51826: gap of 100 bp
* 51827 52542: contig of 716 bp in length
* 52543 52642: gap of 100 bp
* 52643 53376: contig of 734 bp in length
* 53377 53476: gap of 100 bp
* 53477 54205: contig of 729 bp in length
* 54206 54305: gap of 100 bp
* 54306 55049: contig of 744 bp in length
* 55050 55149: gap of 100 bp
* 55150 55854: contig of 705 bp in length
* 55855 55954: gap of 100 bp
* 55955 56709: contig of 755 bp in length

Query Match      83.6%; Score 18.4; DB 14; Length 64091;
Best Local Similarity 95.0%; Pred. No. 1.6e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ATGCTCACACATTAACCTGGA 22
    |||||
Db 32504 ATGCACACACATTAACCTGGA 32523

RESULT 16
AC119858/c
LOCUS      AC119858      80308 bp      DNA      linear      HTG 02-MAY-2002
DEFINITION Mus musculus clone RP23-257B9, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC119858
VERSION   AC119858.1 GI:20389501
KEYWORDS  HTG, HTGS PHASE0.
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
          Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 80308)
AUTHORS   Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE     Mus musculus, clone RP23-257B9
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 80308)
AUTHORS   Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
          Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguski, L.,
          Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
          Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,
          Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
          Fero, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
          Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
          Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
          Kamat, A., Karatas, A., Kells, C., Lakocque, K., Lamazares, R.,
          Landers, T., Lehoczyk, J., Levine, R., Lindblad-Toh, K., Liu, G.,
          MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
          McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L.,
          Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,

```

TITLE  
JOURNAL  
COMMENT

Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D.,  
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,  
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,  
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,  
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,  
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (02-MAY-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L22559  
Center clone name: 257\_B\_9  
-----  
\* NOTE: This record contains 74 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.  
\* 1 985: contig of 985 bp in length  
\* 986 1085: gap of 100 bp  
\* 1086 2085: contig of 1000 bp in length  
\* 2086 2185: gap of 100 bp  
\* 2186 3177: contig of 992 bp in length  
\* 3178 3277: gap of 100 bp  
\* 3278 4240: contig of 963 bp in length  
\* 4241 4340: gap of 100 bp  
\* 4341 5322: contig of 982 bp in length  
\* 5323 5422: gap of 100 bp  
\* 5423 6410: contig of 988 bp in length  
\* 6411 7485: contig of 975 bp in length  
\* 6511 7585: gap of 100 bp  
\* 7586 8530: contig of 945 bp in length  
\* 8531 8631: gap of 100 bp  
\* 8631 9637: contig of 1006 bp in length  
\* 9637 9736: gap of 100 bp  
\* 9737 10742: contig of 1006 bp in length  
\* 10743 10842: gap of 100 bp  
\* 10843 11849: contig of 1007 bp in length  
\* 11850 11949: gap of 100 bp  
\* 11950 12966: contig of 1017 bp in length  
\* 12967 13066: gap of 100 bp  
\* 13067 14095: contig of 1029 bp in length  
\* 14096 14195: gap of 100 bp  
\* 14196 15276: contig of 981 bp in length  
\* 15177 15276: gap of 100 bp  
\* 15277 16288: contig of 852 bp in length  
\* 16289 16228: gap of 100 bp  
\* 16229 17228: contig of 1000 bp in length  
\* 17229 17328: gap of 100 bp  
\* 17329 18328: contig of 1000 bp in length  
\* 18329 18428: gap of 100 bp  
\* 18429 19433: contig of 1005 bp in length  
\* 19434 19533: gap of 100 bp  
\* 19534 20506: contig of 973 bp in length  
\* 20507 20606: gap of 100 bp

\* 20607 21576: contig of 970 bp in length  
\* 21577 21676: gap of 100 bp  
\* 21677 22629: contig of 953 bp in length  
\* 22630 22729: gap of 100 bp  
\* 22730 23708: contig of 979 bp in length  
\* 23709 23808: gap of 100 bp  
\* 23809 24765: contig of 957 bp in length  
\* 24766 24865: gap of 100 bp  
\* 24866 25847: contig of 982 bp in length  
\* 25848 26939: contig of 992 bp in length  
\* 26940 27039: gap of 100 bp  
\* 27040 28021: contig of 982 bp in length  
\* 28022 28121: gap of 100 bp  
\* 28122 29117: contig of 996 bp in length  
\* 29118 29217: gap of 100 bp  
\* 29219 30214: contig of 997 bp in length  
\* 30215 30314: gap of 100 bp  
\* 30315 31314: contig of 1000 bp in length  
\* 31315 31414: gap of 100 bp  
\* 31415 32375: contig of 961 bp in length  
\* 32376 32475: gap of 100 bp  
\* 32476 33443: contig of 968 bp in length  
\* 33444 33543: gap of 100 bp  
\* 33544 34531: contig of 988 bp in length  
\* 34532 34631: gap of 100 bp  
\* 34632 35644: contig of 1013 bp in length  
\* 35645 35744: gap of 100 bp  
\* 35745 36703: contig of 959 bp in length  
\* 36704 37790: gap of 100 bp  
\* 37791 37890: contig of 987 bp in length  
\* 37891 38881: contig of 991 bp in length  
\* 38882 39889: gap of 100 bp  
\* 39890 40889: contig of 1008 bp in length  
\* 40890 41125: gap of 100 bp  
\* 41126 41225: contig of 1036 bp in length  
\* 41226 42240: contig of 1015 bp in length  
\* 42241 42340: gap of 100 bp  
\* 42341 43316: contig of 976 bp in length  
\* 43317 43416: gap of 100 bp  
\* 43417 44396: contig of 980 bp in length  
\* 44397 44496: gap of 100 bp  
\* 44497 45508: contig of 1012 bp in length  
\* 45509 46569: contig of 961 bp in length  
\* 46570 46669: gap of 100 bp  
\* 46670 47664: contig of 995 bp in length  
\* 47665 47764: gap of 100 bp  
\* 47765 48779: contig of 1015 bp in length  
\* 48780 48879: gap of 100 bp  
\* 48880 49863: contig of 984 bp in length  
\* 49864 49963: gap of 100 bp  
\* 49964 50970: contig of 1007 bp in length  
\* 50971 51071: gap of 100 bp  
\* 51071 52065: contig of 995 bp in length  
\* 52066 52165: gap of 100 bp  
\* 52166 53188: contig of 1023 bp in length  
\* 53189 53288: gap of 100 bp  
\* 53289 54243: contig of 955 bp in length  
\* 54244 54343: gap of 100 bp  
\* 54344 55346: contig of 1003 bp in length  
\* 55347 55446: gap of 100 bp  
\* 55447 56431: contig of 985 bp in length  
\* 56432 56531: gap of 100 bp  
\* 56532 57522: contig of 991 bp in length  
\* 57523 58626: contig of 1004 bp in length  
\* 58627 58727: gap of 100 bp  
\* 58728 59827: contig of 1001 bp in length  
\* 59828 60838: contig of 1011 bp in length

\* 60839 60938: gap of 100 bp  
\* 60939 61930: contig of 992 bp in length  
\* 61931 62030: gap of 100 bp  
\* 62031 62990: contig of 960 bp in length  
\* 62991 63090: gap of 100 bp  
\* 63091 64039: contig of 949 bp in length  
\* 64040 64139: gap of 100 bp  
\* 64140 65111: contig of 972 bp in length  
\* 65112 66157: contig of 946 bp in length  
\* 66158 66257: gap of 100 bp  
\* 66258 67279: contig of 1022 bp in length  
\* 67280 67379: gap of 100 bp  
\* 67380 68380: contig of 1001 bp in length  
\* 68381 69512: contig of 1032 bp in length  
\* 69513 70626: contig of 1014 bp in length  
\* 70627 70726: gap of 100 bp  
\* 70727 71727: contig of 1001 bp in length  
\* 71728 72813: contig of 986 bp in length  
\* 72814 72913: gap of 100 bp  
\* 72914 73865: contig of 952 bp in length  
\* 73866 73965: gap of 100 bp  
\* 73966 74915: contig of 950 bp in length

Query Match 83.6%; Score 18.4; DB 14; Length 80308;  
Best Local Similarity 95.0%; Pred. No. 1.6e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCATGCTCACATTAACTG 20  
|||||||  
Db 49213 TCATGCTCACAGTTAACTG 49194

RESULT 17  
AC147104/c

LOCUS 168370 bp DNA linear ROD 29-MAY-2004  
DEFINITION Mus musculus BAC clone RP23-466J20 from chromosome 14, complete sequence.  
ACCESSION AC147104  
VERSION AC147104.2 GI:47777601  
KEYWORDS HTG.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 168370)  
Levy, A., Bielicki, L. and Haakenson, W.  
TITLE The sequence of Mus musculus BAC clone RP23-466J20  
JOURNAL Unpublished (2001)  
AUTHORS Wilson, R.K.  
REFERENCE 2 (bases 1 to 168370)  
Direct Submission  
TITLE Direct Submission  
JOURNAL Submitted (04-NOV-2003) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
REFERENCE 3 (bases 1 to 168370)  
Wilson, R.K.  
TITLE Direct Submission  
JOURNAL Submitted (28-MAY-2004) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
REFERENCE 4 (bases 1 to 168370)  
Wilson, R.K.  
TITLE Direct Submission  
JOURNAL Submitted (29-MAY-2004) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
COMMENT On May 28, 2004 this sequence version replaced gi:38154165.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC

Web site: <http://genome.wustl.edu>  
 Contact: [submissions@watson.wustl.edu](mailto:submissions@watson.wustl.edu)  
 ----- Summary Statistics -----  
 ----- Center project name: M\_BA0466J20 -----

## NOTICE:

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

## MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

## SOURCE INFORMATION:

The RPCI-23 BAC library has been constructed by Kazutoyo Osegawa and Minako Tateno in the laboratory of Pieter de Jong (<http://www.chori.org>) from female C57BL/6J mouse kidney and/or brain genomic DNA. The clone and detailed information can be obtained from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

## NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone. This clone is overlapped by AC126258 and AC123945.

## FEATURES

## source

```

1. .168370
   /organism="Mus musculus"
   /mol_type="genomic DNA"
   /db_xref="taxon:10090"
   /chromosome="14"
   /map="14"
   /clone="RP23-466J20"
   /clone_lib="RPCI-23"
   repeat_region      284..706
   /rpt_family="MaLR"
   repeat_region      1384..1465
   /rpt_family="B4"
   repeat_region      1812..1886
   /rpt_family="L2"
   repeat_region      1887..1995
   /rpt_family="MER1_type"
   repeat_region      2242..2636
   /rpt_family="MaLR"
   repeat_region      4523..4671
   /rpt_family="Alu"
   repeat_region      4749..4932
   /rpt_family="ERV1"
   repeat_region      5087..5361
   /rpt_family="B4"
   repeat_region      6908..7015
   /rpt_family="Alu"
   repeat_region      8522..8736
   /rpt_family="B2"
   repeat_region      9655..9731
   /rpt_family="Alu"
   repeat_region      11092..11307
   /rpt_family="B2"
   repeat_region      12130..12487
   /rpt_family="L1"
   repeat_region      12566..12758
   /rpt_family="L1"
   repeat_region      13710..14088
   /rpt_family="L1"
   repeat_region      14787..15241
   /rpt_family="L1"

/rpt_family="RMR15"
1752..17827
/rpt_family="MIR"
18070..18321
/rpt_family="MaLR"
20080..20626
/rpt_family="L1"
20717..20896
/rpt_family="L1"
20911..21153
/rpt_family="L2"
21254..21466
/rpt_family="B4"
21467..21746
/rpt_family="B4"
22002..22106
/rpt_family="L2"
22190..22346
/rpt_family="MIR"
22492..23262
/rpt_family="L1"
23259..24129
/rpt_family="L1"
24290..24348
/rpt_family="ERV1"
24993..25359
/rpt_family="L2"
25560..25618
/rpt_family="ERV1"
27907..27960
/rpt_family="ERV1"
28447..28642
/rpt_family="B2"
29442..29713
/rpt_family="B4"
34294..34467
/note="Sequence derived from one plasmid subclone."
35151..35415
/note="Sequence derived from PCR product of genomic DNA."
37118..37511
/rpt_family="MaLR"
37723..38124
/rpt_family="MaLR"
38295..38649
/rpt_family="MaLR"
39062..39158
/rpt_family="CRI"
39727..39840
/rpt_family="ERVK"
39841..40112
/rpt_family="B4"
41322..41491
/rpt_family="MaLR"
41474..41600
/rpt_family="ERVK"
41601..41806
/rpt_family="MaLR"
41807..41820
/rpt_family="ERVK"
42617..44726
/rpt_family="L1"
44982..45353
/rpt_family="L1"
45354..45402
/rpt_family="L1"
45409..46204
/rpt_family="L1"
46255..48179
/rpt_family="L1"
48308..48781
/rpt_family="L1"
50401..50914
/rpt_family="L1"

```

```

repeat_region 50610. .51008
/rpt_family="L1MB4_SLINE/L1"
repeat_region 51082. .51295
/rpt_family="L1"
repeat_region 51390. .51892
/rpt_family="L1"
repeat_region 52064. .52241
/rpt_family="L1"
repeat_region 52253. .52461
/rpt_family="L1"
repeat_region 52474. .52663
/rpt_family="L1"
repeat_region 52858. .53212
/rpt_family="L1"
repeat_region 53842. .54305
/rpt_family="PMER6A"
repeat_region 54324. .54815
/rpt_family="L1"
repeat_region 56004. .56150
Query Match 83.6%; Score 18.4; DB 9; Length 168370;
Best Local Similarity 95.0%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 3 ATGCTACACATTAACCTGGA 22
|||||
Db 130868 ATGCACACATTAACCTGGA 130849

```

```

RESULT 18
AC163486/c
LOCUS AC163486 192120 bp DNA linear HTG 12-JUN-2005
DEFINITION Mus musculus chromosome 1 clone RP23-257C12 map 1, WORKING DRAFT
SEQUENCE 28 unordered pieces.
AC163486
AC163486.1 GI:67514684
HTG; HTGS_PHASE1; HTGS DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
Birren,B., Nusbaum,C. and Lander,E.
1 (bases 1 to 192120)
Mus musculus chromosome 1, clone RP23-257C12
Unpublished
2 (bases 1 to 192120)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
Bloom,T., Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J.,
Choepe,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,
DeArillano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
Erickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D.,
Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,
Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,
McCarthy,M., Meldrum,J., Meneus,L., Minova,T., Mienga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupbach,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (12-JUN-2005) Broad Institute of MIT and Harvard, 320
Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)

```

```

http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Broad Institute of MIT and Harvard
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@broad.mit.edu
----- Project Information
Center project name: L3280
Center clone name: 257_C12
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 182161 bases at least Q40
Consensus quality: 185301 bases at least Q30
Consensus quality: 187649 bases at least Q20
Insert size: 230000; agarose-fp
Insert size: 189420; sum-of-contigs
Quality coverage: 3.7 in Q20 bases; agarose-fp
Quality coverage: 4.5 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 227: contig of 227 bp in length
* 327: gap of unknown length
* 328
* 1250: contig of 923 bp in length
* 1251
* 1350: gap of unknown length
* 1351
* 1430: contig of 80 bp in length
* 1431
* 1530: gap of unknown length
* 1531
* 2386: contig of 856 bp in length
* 2387
* 2487
* 3698: contig of 1212 bp in length
* 3699
* 3798: gap of unknown length
* 3799
* 4993: contig of 1195 bp in length
* 4994
* 5093: gap of unknown length
* 5094
* 5864: contig of 1771 bp in length
* 5865
* 6965: contig of 1810 bp in length
* 6966
* 8774: gap of unknown length
* 8775
* 8875
* 41598: contig of 32724 bp in length
* 41599
* 41599
* 44599: contig of 2901 bp in length
* 44599
* 44599: gap of unknown length
* 44600
* 47337: contig of 2638 bp in length
* 47338
* 47337: gap of unknown length
* 47338
* 50132: contig of 2695 bp in length
* 50133
* 50232: gap of unknown length
* 50233
* 52300: contig of 2068 bp in length
* 52301
* 52400: gap of unknown length
* 52401
* 56635: contig of 4235 bp in length
* 56636
* 56735: gap of unknown length
* 56736
* 60231: contig of 3496 bp in length
* 60232
* 60331: gap of unknown length
* 60332
* 64327: contig of 3996 bp in length
* 64328
* 64427: gap of unknown length
* 64428
* 69082: contig of 4655 bp in length
* 69083
* 69182: gap of unknown length
* 69183
* 73825: contig of 4643 bp in length
* 73826
* 73925: gap of unknown length
* 73926
* 78758: contig of 4833 bp in length
* 78759
* 78858: gap of unknown length
* 78859
* 84461: contig of 5603 bp in length
* 84462
* 84561: gap of unknown length
* 84562
* 89277: contig of 4716 bp in length
* 89278
* 89377: gap of unknown length
* 89378
* 95743: contig of 6366 bp in length
* 95744
* 95843: gap of unknown length

```

\* 95844 103913: contig of 8070 bp in length  
 \* 103914 104013: gap of unknown length  
 \* 104014 116629: contig of 12616 bp in length  
 \* 116629 116729: gap of unknown length  
 \* 116730 134098: contig of 17369 bp in length  
 \* 134099 134198: gap of unknown length  
 \* 134199 165328: contig of 31130 bp in length  
 \* 165329 165428: gap of unknown length  
 \* 165429 190983: contig of 25555 bp in length  
 \* 190984 191084: gap of unknown length  
 \* 191084 192120: contig of 1037 bp in length.

## FEATURES

Location/Qualifiers  
 1..192120  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosomes="1"  
 /map="1"  
 /clone="RP23-257C12"  
 /clone\_lib="RPCI-23 Female Mouse BAC"

misc\_feature 1..227  
 /notes="assembly\_fragment  
 clone\_end:SP6  
 vector\_side:left"

gap 228..327

misc\_feature 328..1250  
 /estimated\_length=unknown

gap 1251..1350

misc\_feature 1351..1430  
 /note="assembly\_fragment"

gap 1431..1530

misc\_feature 1531..2236  
 /estimated\_length=unknown

gap 2237..2486

misc\_feature 2487..3698  
 /note="assembly\_fragment"

gap 3699..3798

misc\_feature 3799..4993  
 /note="assembly\_fragment"

gap 4994..5093

misc\_feature 5094..6864  
 /note="assembly\_fragment"

gap 6865..6964

misc\_feature 6965..8774  
 /note="assembly\_fragment"

gap 8775..8874

misc\_feature 8875..41598  
 /note="assembly\_fragment"

gap 41599..41698

misc\_feature 41699..44599  
 /note="assembly\_fragment"

gap 44600..44699

misc\_feature 44700..47337  
 /estimated\_length=unknown

gap 47338..47437

misc\_feature 47438..50132  
 /note="assembly\_fragment"

gap 50133..50232

misc\_feature 50233..52300  
 /note="assembly\_fragment"

gap 52301..52400

misc\_feature /estimated\_length=unknown  
 52401..56635  
 /note="assembly\_fragment"  
 gap 56636..56735  
 /estimated\_length=unknown  
 misc\_feature 56736..60231  
 /note="assembly\_fragment"  
 gap 60232..60331  
 /estimated\_length=unknown  
 misc\_feature 60332..64327  
 /note="assembly\_fragment"  
 gap 64328..64427

Query Match 83.6%; Score 18.4; DB 14; Length 192120;  
 Best Local Similarity 95.0%; Pred. No. 1.5e+02;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCATGCTCACACATTAACTG 20  
 ||||| ||||| ||||| |||||

Db 16595 TCATGCTCACACATTAACTG 16576

## RESULT 19

AC126258

LOCUS

DEFINITION

AC126258

VERSION

KEYWORDS

SOURCE

ORGANISM

AC126258

HTG.

Mus musculus

Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridea; Muridae; Murinae; Mus.

Goyea, E. and Meyer, R.

1 (bases 1 to 203304)

The sequence of Mus musculus BAC clone RP23-325B8

Unpublished (2001)

2 (bases 1 to 203304)

Wilson, R.

Sequencing of Mus musculus

Unpublished (2001)

3 (bases 1 to 203304)

McPherson, J.D. and Waterston, R.H.

Direct Submission

Submitted (04-JUL-2002) Genome Sequencing Center, 4444 Forest Park

Parkway, St. Louis, MO 63108, USA

4 (bases 1 to 203304)

McPherson, J.D. and Waterston, R.H.

Direct Submission

Submitted (14-APR-2003) Genome Sequencing Center, 4444 Forest Park

Parkway, St. Louis, MO 63108, USA

5 (bases 1 to 203304)

Wilson, R.K.

Direct Submission

Submitted (22-MAY-2003) Genome Sequencing Center, 4444 Forest Park

Parkway, St. Louis, MO 63108, USA

6 (bases 1 to 203304)

Wilson, R.

Direct Submission

Submitted (25-NOV-2003) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

On May 22, 2003 this sequence version replaced gi:29824559.

----- Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu>

Contact: [submissions@watson.wustl.edu](mailto:submissions@watson.wustl.edu)

----- Summary Statistics

-----

Center project name: M\_BA0325B08



NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

#### SOURCE INFORMATION:

The RPCI-23 BAC Library has been constructed by Kazutoyo Osegawa and Minako Tateno in the laboratory of Pieter de Jong (<http://www.chori.org>) from female C57BL/6J mouse kidney and/or brain genomic DNA. The clone and detailed information can be obtained from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

#### NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone.

#### FEATURES

repeat_region	23847..24194	/rpt_family="MaLR"
repeat_region	26678..26925	/rpt_family="MaLR"
repeat_region	27050..27158	/rpt_family="B4"
repeat_region	27586..27647	/rpt_family="MER2_type"
repeat_region	28642..28983	/rpt_family="ID"
repeat_region	29069..29125	/rpt_family="MaLR"
repeat_region	29085..29325	/rpt_family="ERV1"
repeat_region	29745..30018	/rpt_family="MaLR"
repeat_region	31336..31415	/rpt_family="B4"
repeat_region	31448..31796	/rpt_family="MaLR"
repeat_region	31797..32192	/rpt_family="MaLR"
repeat_region	32193..33287	/rpt_family="MaLR"
repeat_region	33288..33683	/rpt_family="MaLR"
repeat_region	33684..33719	/rpt_family="MaLR"
repeat_region	35452..35807	/rpt_family="MaLR"
repeat_region	35870..36104	/rpt_family="ERVK"
repeat_region	36411..36499	/rpt_family="ERVK"
repeat_region	36486..36594	/rpt_family="B2"
repeat_region	36702..36804	/rpt_family="B2"
repeat_region	38043..38178	/rpt_family="L1"
repeat_region	40107..40489	/rpt_family="Alu"
repeat_region	40605..41038	/rpt_family="ERVK"
repeat_region	41354..41568	/rpt_family="MaLR"
repeat_region	42323..42381	/rpt_family="B2"
repeat_region	42431..42724	/rpt_family="MaLR"
repeat_region	46485..46619	/rpt_family="MaLR"
repeat_region	47001..47107	/rpt_family="MaLR"
repeat_region	47157..47259	/rpt_family="L1"
repeat_region	47921..48093	/rpt_family="MaLR"
repeat_region	51165..51298	/rpt_family="MER2_type"
repeat_region	53161..53659	/rpt_family="Alu"
repeat_region	56110..56442	/rpt_family="MaLR"
repeat_region	58667..58723	/rpt_family="MaLR"
repeat_region	59170..59243	/rpt_family="Alu"
repeat_region	59365..59854	/rpt_family="L1"
repeat_region	61891..62005	/rpt_family="ERVK"
repeat_region		/rpt_family="MaLR"

repeat\_region 63945..64049  
 /rpt\_family="L1"  
 repeat\_region 64073..64179  
 /rpt\_family="B2"  
 repeat\_region 64364..64513

Query Match 83.6%; Score 18.4; DB 9; Length 203304;

Best Local Similarity 95.0%; Pred. No. 1.5e+02;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ATGCTCACATTAAGTGA 22

DB 153244 ATGCACACATTAAGTGA 153263

RESULT 20

AC153881/C

LOCUS

DEFINITION AC153881 208992 bp DNA linear ROD 29-MAY-2005  
 Mus musculus 10 BAC RP23-257B9 (Roswell Park Cancer Institute  
 (C57BL/6J Female) Mouse BAC Library) complete sequence.

ACCESSION AC153881

VERSION AC153881.4 GI:58372173

KEYWORDS HTG.

SOURCE Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS

Muzny, D., Adams, C., Agbai, II, O., Allen, C., Alsbrooks, S., Archer, P.,  
 Arradondo, H., Bandaranaike, D., Bangura, L., Beltran, B., Beltran, R.,  
 Beraducci, A., Biswal, K., Blyth, P., Bonham, H., Buhay, C., Burch, P.,  
 Cadoree, I., Canada, A., Cardenas, V., Carter, K., Cavazos, I.,  
 Chacko, J., Chahrour, M., Chavez, D., Chen, A., Chen, G., Chen, R.,  
 Cheng, M.-T., Chu, J., Clerc, K., Cockrell, R., Coyle, M., Cree, A.,  
 Curry, S., Dai, W., Davila, M.L., Davis, C., Davy-Carroll, L., De  
 Anda, C., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H.,  
 Donlin, J., McCauley, S., Dugan-Rocha, S., Dunn, A., Durbin, K.,  
 Dziuda, D., Egan, A., Escotto, M., Espinosa, V., Eugene, C., Fa, M.,  
 Fernandez, S., Fernando, P., Flagg, N., Forbes, D., Foster, P.,  
 Fowler, G., Fu, Q., Fuh, E., Garcia, A., Garcia, R., Garner, T.,  
 Gaskin, C., Gench, S., Ghose, S., Gill, R., Gonzalez, D.,  
 Gonzalez-Garay, M., Guevara, W., Holder, M., Holders, W., Haeberlein, K.,  
 Hall, B., Hanid, H., Hamilton, K., Harbes, B., Harris, R., Havlak, P.,  
 Hawes, A., Hawkins, E., Hayes, S., Hemphill, D., Hernandez, J.,  
 Hines, S., Hitchens, M., Hodgson, A., Hogue, M., Hollins, B.,  
 Howell, L.T., Hulyk, S., Hume, J., Imo, K., Jackson, A., Jackson, L.,  
 Jacob, L., Jiang, H., Johnson, B., Johnson, R., Kalafus, K., Kelly, S.,  
 Keys, T., Khan, Z., King, L., Kovar, C., Kowis, A., Kowis, C., Lara, F.,  
 Leal, S., Lee, K., Lee, S., Legall, F.I., Lemon, S., Lewis, L., Li, B.,  
 Li, Y., Li, Z., Linnell, M., Liu, W., Liu, Y.-S., Liu, Y., Liyanage, D.,  
 London, P., Lopez, J., Lorensu, L., Lozano, R., Luk, T., Madu, R.,  
 Maheshwari, M., Mahoney, C., Malloy, K., Mansouri, D., Martinez, E.,  
 McCrelland, H., McPherson, J., Mercadao, C., Metzker, M.,  
 Milosavljevic, A., Minja, E., Morgan, M., Morris, S., Munidasa, M.,  
 Murray, D., Nazareth, L., Ngo, D., Nguyen, N., Norwig-Bastaugh, E.,  
 Nott, A., Nwaokiemhen, O., Oregon, M., Ochi-Okorie, C., Oden, E.,  
 Okwuonu, G., Okwuonu, K., Parker, D., Pasternak, S., Patel, B.,  
 Patel, V., Paul, H., Perez, A., Perez, L., Petrosino, J., Pham, T.,  
 Primus, E., Pu, L.-L., Puazo, M., Qin, X., Quinn, A., Quiroz, J.,  
 Rabata, D., Rachlin, E., Reigh, R., Ren, Y., Reuter, M., Richards, S.,  
 Rives, C., Rodriguez, F., Rojao, A., Ruiz, S.J., Sana, M., Sanders, W.,  
 Santibanez, J., Santos, R., Savery, G., Scherer, S., Shen, H., Shen, Y.,  
 Sisson, I., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R.,  
 Svatek, A., Taylor, E., Taylor, T., Thomas, N., Thorn, R., Thornton, R.,  
 Trejos, Z., Usmami, K., Vargu, C., Verduco, D., Villaseana, D., Virk, D.,  
 Volkov, A., Waldron, L., Walker, B., Wang, Q., Wang, S., Warren, J.,  
 Wei, X., Wheeler, D., Williams, G., Williams, R., Worley, K., Wright, R.,  
 Wu, J., Yakub, S., Yan, K., Yuan, Y., Yu, F., Zhang, J., Zhang, L.,  
 Zhang, Z., Zhou, J., Weinstock, G. and Gibbs, R.

TITLE

JOURNAL

REFERENCE 2 (bases 1 to 208992)

AUTHORS

TITLE

JOURNAL

Worley, K.C.  
 Direct Submission  
 Submitted (18-DEC-2004) Human Genome Sequencing Center, Baylor  
 College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 208992)

REFERENCE

TITLE

JOURNAL

Worley, K.C.  
 Direct Submission  
 Submitted (22-JAN-2005) Human Genome Sequencing Center, Baylor  
 College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 4 (bases 1 to 208992)

REFERENCE

TITLE

JOURNAL

Worley, K.C.  
 Direct Submission  
 Submitted (01-FEB-2005) Human Genome Sequencing Center, Baylor  
 College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 5 (bases 1 to 208992)

REFERENCE

TITLE

JOURNAL

Worley, K.C.  
 Direct Submission  
 Submitted (29-MAY-2005) Human Genome Sequencing Center, Baylor  
 College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 On Feb 1, 2005 this sequence version replaced gi:58036505.

COMMENT

Sequencing is completed to a minimum standard of double strand  
 coverage with a minimum of 2 clones and 2 reads with no ambiguities  
 or 2 chemistries with a minimum of 2 clones and 3 reads with no  
 ambiguities. If the sequence quality does not meet this standard,  
 it will be indicated in the annotation.

The repeat regions shown were identified using RepeatMasker by  
 Adrian Smit.

Sequence similarities were identified using Powerblast by Jinghui  
 Zhang.

Exon/Intron boundaries of identified genes were chosen if there  
 were canonical splice junctions that maintained sequence continuity  
 across the splice junctions.

FEATURES

source

1..208992  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /clone="RP23-257B9"  
 /complement(1..38844)  
 /note="overlaps bases 1..38844 of clone AC158598"  
 /function="clone overlap"  
 complement(421..788)  
 /rpt\_family="L1\_Mm"

misc\_feature

repeat\_region  
 complement(421..788)  
 /rpt\_family="L1\_Mm"

repeat\_region

repeat\_region  
 complement(948..1310)  
 /rpt\_family="L1Md\_F3"

repeat\_region

repeat\_region  
 complement(1475..1787)  
 /rpt\_family="MLT2B3"

repeat\_region

repeat\_region  
 complement(1..38844)  
 /rpt\_family="TATAA)n"

repeat\_region

repeat\_region  
 complement(3136..3298)  
 /rpt\_family="RSINE1"

repeat\_region

repeat\_region  
 complement(4422..4583)  
 /rpt\_family="B3"

repeat\_region

repeat\_region  
 complement(4651..4666)  
 /rpt\_family="Lx3A"

repeat\_region

repeat\_region  
 complement(4697..5243)  
 /rpt\_family="Lx3A"

repeat\_region

repeat\_region  
 complement(5244..5505)  
 /rpt\_family="Lx3A"

repeat\_region

repeat\_region  
 complement(5814..5993)

repeat_region	/rpt_family="Lx9"	Best Local Similarity 95.0%; Pred. No. 1.5e+02;			
repeat_region	6082. .6104	Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;			
repeat_region	/rpt_family="AT_rich"				
repeat_region	6241. .6296				
repeat_region	/rpt_family="(TATATG)n"				
repeat_region	complement(6467. .6895)				
repeat_region	/rpt_family="RMER17A2"				
repeat_region	complement(6896. .7135)				
repeat_region	/rpt_family="L1_Mus1"				
repeat_region	complement(7136. .7679)				
repeat_region	/rpt_family="RMER17A2"				
repeat_region	1011. .11250				
repeat_region	/rpt_family="L1Md_T"				
repeat_region	11261. .11365				
repeat_region	/rpt_family="L1_Mur1"				
repeat_region	11358. .13005				
repeat_region	/rpt_family="Lx8"				
repeat_region	13891. .13913				
repeat_region	/rpt_family="AT_rich"				
repeat_region	13918. .14023				
repeat_region	/rpt_family="B1_Mus1"				
repeat_region	14025. .14137				
repeat_region	/rpt_family="(TA)n"				
repeat_region	14139. .14477				
repeat_region	/rpt_family="(TACA)n"				
repeat_region	complement(15043. .15388)				
repeat_region	/rpt_family="ORR1A0"				
misc_feature	15663				
	/notes="IS2 element removed;"				
repeat_region	/functions="bacterial transposon excised"				
repeat_region	15663				
repeat_region	/rpt_family="IS2"				
repeat_region	15668. .15731				
repeat_region	/rpt_family="(TC)n"				
repeat_region	15732. .15760				
repeat_region	/rpt_family="(CA)n"				
repeat_region	15814. .16090				
repeat_region	/rpt_family="Lx6"				
repeat_region	16100. .16278				
repeat_region	/rpt_family="(GAAA)n"				
repeat_region	16279. .16437				
repeat_region	/rpt_family="(GA)n"				
repeat_region	16445. .17169				
repeat_region	/rpt_family="Lx6"				
repeat_region	17347. .17895				
repeat_region	/rpt_family="Lx6"				
repeat_region	18148. .18199				
repeat_region	/rpt_family="AT_rich"				
repeat_region	18376. .18436				
repeat_region	/rpt_family="(CATATA)n"				
repeat_region	19464. .19500				
repeat_region	/rpt_family="(A)n"				
repeat_region	19968. .20003				
repeat_region	/rpt_family="(CAT)n"				
repeat_region	complement(20012. .20652)				
repeat_region	/rpt_family="Lx3B"				
repeat_region	complement(20653. .21085)				
repeat_region	/rpt_family="RMER17D2"				
repeat_region	21086. .21259				
repeat_region	/rpt_family="(GGAGA)n"				
repeat_region	complement(21260. .21580)				
repeat_region	/rpt_family="RMER17D2"				
repeat_region	complement(21581. .21783)				
repeat_region	/rpt_family="Lx3B"				
repeat_region	22066. .22423				
repeat_region	/rpt_family="RLTR9A"				
repeat_region	complement(22510. .22894)				
repeat_region	/rpt_family="L1"				
repeat_region	complement(23039. .23636)				
repeat_region	/rpt_family="L1_Mur3"				
repeat_region	23667. .25365				
Query Match	83.6%; Score 18.4; DB 9; Length 208992;				

Best Local Similarity 95.0%; Pred. No. 1.5e+02;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCATGCTCACACATTAACTG 20

Db 108850 TCATGCTCACAGTTAACTG 108831

RESULT 21

AC159476

LOCUS

DEFINITION Mus musculus 10 BAC RP23-53016 (Roswell Park Cancer Institute (C57BL/6J Female) Mouse BAC Library) complete sequence.

AC159476

AC159476.5 GI:66793494

HTG. |||||||

Mus musculus (house mouse)

Mus musculus

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 212483)

Muzny, D., Adams, C., Agbai II, O., Allen, C., Alsbrooks, S., Archer, P., Arraondo, H., Bandaranaike, D., Bangura, L., Beltran, B., Beltran, R., Beraducci, A., Biswal, K., Blyth, P., Bonham, H., Buhay, C., Burch, P., Cadoree, I., Canada, A., Cardenas, V., Carter, K., Cavazos, I., Chacko, J., Chahrouh, M., Chavez, D., Chen, A., Chen, G., Chen, R., Cheng, M.-T., Chu, J., Clerc, K., Cockrell, R., Coyle, M., Cree, A., Curry, S., Dai, W., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Donlin, J., McCauley, S., Dugan-Rocha, S., Dunn, A., Durbin, K., Dziuda, D., Egan, A., Escotto, M., Espinosa, V., Eugene, C., Fa, M., Fernandez, S., Fernando, P., Flagg, N., Forbes, L., Foster, P., Fowler, G., Fu, Q., Fuh, E., Garcia, A., Garcia, R., Garner, T., Gaskin, C., Gench, S., Ghose, S., Gill, R., Gonzalez, D., Gonzalez-Garay, M., Guevara, W., Holder, M., Haaland, W., Haerlen, K., Hall, B., Hamid, H., Hamilton, K., Harbes, B., Harris, R., Havlak, P., Hawes, A., Hawkins, E., Hayes, S., Hemphill, L., Hernandez, J., Hines, S., Hitchens, M., Hodgson, A., Hogues, M., Hollins, B., Howell, L. T., Hulyk, S., Hume, J., Imo, K., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Kalafus, K., Kelly, S., Keys, T., Khan, Z., King, L., Kovar, C., Kowis, A., Kowis, C., Lara, F., Leal, S., Lee, K., Lee, S., LeCall, F. I., Lemon, S., Lewis, L., Li, B., Li, Y., Li, Z., Linnell, M., Liu, W., Liu, Y.-S., Liu, Y., Liyanage, D., London, P., Lopez, J., Lorensuhewa, L., Lozano, R., Luk, T., Madu, R., Maheshwari, M., Mahoney, C., Malloy, K., Mansouri, D., Martinez, E., McClelland, H., McPherson, J., Mercadao, C., Metzker, M., Milosavljevic, A., Minja, E., Morgan, M., Morris, S., Munidasa, M., Murray, D., Nazareth, L., Ngo, D., Nguyen, N., Norwig-Eastaugh, E., Nott, A., Nwaokeme, O., Obregon, M., Ochi-Okorie, C., Odeh, E., Okwuonu, G., Okwuonu, K., Parker, D., Pasternak, S., Patel, B., Patel, V., Paul, H., Perez, A., Perez, L., Petrosino, J., Pham, T., Primus, E., Pu, L.-L., Puazo, M., Qin, X., Quinn, A., Quiros, J., Rabata, D., Rachlin, E., Reigh, R., Ren, Y., Reuter, M., Richards, S., Rivers, C., Rodriguez, F., Rojas, A., Ruiz, S. J., Sana, M., Sanders, W., Santibanez, J., Santos, R., Savery, G., Scherer, S., Shen, H., Shen, Y., Sisson, I., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Svatek, A., Taylor, E., Taylor, T., Thomas, N., Thorn, R., Thornton, R., Trejos, Z., Usmani, K., Vargo, C., Verdusco, D., Villabana, D., Virk, D., Volkov, A., Waldron, L., Walker, B., Wang, Q., Wang, S., Warren, J., Wei, X., Wheeler, D., Williams, G., Williams, R., Worley, K., Wright, R., Wu, J., Yakub, S., Yan, K., Yuan, Y., Yu, F., Zhang, J., Zhang, L., Zhang, Z., Zhou, J., Weinstock, G. and Gibbs, R.

Direct Submission

Unpublished

2 (bases 1 to 212483)

Worley, K.C.

Direct Submission

Submitted (04-APR-2005) Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 212483)

Worley, K.C.

Direct Submission

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL Submitted (18-MAY-2005) Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 4 (bases 1 to 212483)

AUTHORS Worley,K.C.

TITLE Direct Submission

JOURNAL Submitted (29-MAY-2005) Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT On May 29, 2005 this sequence version replaced gi:66267755. Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality does not meet this standard, it will be indicated in the annotation.

The repeat regions shown were identified using RepeatMasker by Adrian Smit.

Sequence similarities were identified using Powerblast by Jinghui Zhang.

Exon/Intron boundaries of identified genes were chosen if there were canonical splice junctions that maintained sequence continuity across the splice junctions.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Drafting Center Code: BCM  
Contact: hgsc-help@bcm.tmc.edu.  
Location/Qualifiers

source 1..212483  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosomes="10"  
/clone="RP23-53016"

misc\_feature 1..46709  
/note="overlaps bases 156592..203296 of clone AC113305"

repeat\_region 1..155  
/function="clone overlap"

repeat\_region 645..807  
/rpt\_family="Lx"

repeat\_region 1361..1382  
/rpt\_family="RLTR21"

repeat\_region 1361..1382  
/rpt\_family="(T)n"

repeat\_region complement(1393..1545)  
/rpt\_family="RLTR188"

repeat\_region 1547..1637  
/rpt\_family="MT2B"

repeat\_region 1716..2028  
/rpt\_family="L1MC3"

repeat\_region 1835..2273  
/rpt\_family="L1MC4"

repeat\_region 2326..2417  
/rpt\_family="L1MC4a"

repeat\_region 3001..3094  
/rpt\_family="(CAGAGA)n"

repeat\_region complement(3097..3429)  
/rpt\_family="IAPLTR1a\_Mm"

repeat\_region 3430..3600  
/rpt\_family="(CAGAGA)n"

repeat\_region complement(4123..4182)  
/rpt\_family="ORR1C1"

repeat\_region complement(4206..4372)  
/rpt\_family="Lx5"

repeat\_region complement(4373..4768)  
/rpt\_family="MTA\_Mm"

repeat\_region complement(4769..5865)  
/rpt\_family="MTA\_Mm-int"

repeat\_region complement(5866..6256)  
/rpt\_family="MTA\_Mm"

repeat\_region complement(6257..8271)  
/rpt\_family="Lx5"

repeat\_region 8274..8334  
/rpt\_family="ORR1C1"

repeat\_region 8679..9657  
/rpt\_family="L1\_Mur2"

repeat\_region 9658..9704  
/rpt\_family="(CA)n"

repeat\_region 9705..10580  
/rpt\_family="L1\_Mur2"

repeat\_region 10581..10723  
/rpt\_family="B1\_Mus1"

repeat\_region 10724..11514  
/rpt\_family="L1\_Mur2"

repeat\_region complement(11507..11702)  
/rpt\_family="L1\_Mus1"

repeat\_region complement(11701..12410)  
/rpt\_family="L1\_Mm"

repeat\_region 12426..12804  
/rpt\_family="L1\_Mus1"

repeat\_region 12811..12931  
/rpt\_family="L1\_Mur3"

repeat\_region complement(12957..13563)  
/rpt\_family="L1\_Mur3"

repeat\_region 13565..14420  
/rpt\_family="RMER13B"

repeat\_region complement(14433..14512)  
/rpt\_family="L1\_Mur3"

repeat\_region 14507..14704  
/rpt\_family="L1\_Mus3"

repeat\_region 14715..15100  
/rpt\_family="Lx"

repeat\_region complement(15110..15943)  
/rpt\_family="L1\_Mur3"

misc\_feature 15944..15993  
/note="variable size of repeats (poly AT): size 22-25, deletion 15944-15949"

repeat\_region 15944..15993  
/function="variable size of repeat between subclones"

repeat\_region 15944..15993  
/rpt\_family="(TA)n"

repeat\_region complement(15994..16078)  
/rpt\_family="L1\_Mur3"

repeat\_region complement(16077..16200)  
/rpt\_family="L1\_Mur3"

repeat\_region complement(16201..16585)  
/rpt\_family="ORR1A3-int"

repeat\_region complement(16608..18072)  
/rpt\_family="RLTR19C-int"

repeat\_region complement(18084..18170)  
/rpt\_family="RLTR19C-int"

repeat\_region 18371..18395  
/rpt\_family="(TA)n"

repeat\_region complement(18396..18541)  
/rpt\_family="B1\_Mus2"

repeat\_region complement(18560..18807)  
/rpt\_family="RLTR19C-int"

repeat\_region 18812..18862  
/rpt\_family="(TA)n"

repeat\_region complement(18881..19204)  
/rpt\_family="IAPEY2\_LTR"

repeat\_region 19218..19261  
/rpt\_family="(TA)n"

repeat\_region complement(19267..19447)  
/rpt\_family="B2\_Mm2"

repeat\_region complement(19449..20270)  
/rpt\_family="RLTR19C-int"

Query Match 83.6%; Score 18.4; DB 9; Length 212483;  
Best Local Similarity 95.0%; Pred. No. 1.5e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCATGCTCACATTAACGTG 20  
|||||  
DB 201011 TCATGCTCACATTAACGTG 201030

## RESULT 22

## AC140497

## LOCUS

## DEFINITION

AC140497 214192 bp DNA linear HTG 14-JUN-2003  
Mus musculus chromosome 10 clone RP24-212K2 map 10, WORKING DRAFT  
SEQUENCE, 9 unordered pieces.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

AC140497 GI:31745361  
HTG; HTGS PHASE1; HTGS DRAFT.  
Mus musculus (house mouse)

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidae; Muridae; Murinae; Mus.

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

Birren, B., Nusbaum, C. and Lander, E.  
1 (bases 1 to 214192)  
Mus musculus chromosome 10, clone RP24-212K2  
Unpublished  
2 (bases 1 to 214192)

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,  
Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,  
Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y.,  
Collamore, A., Cook, A., Cooke, P., Corum, B., DeArellano, K.,  
Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,  
Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,  
Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,  
Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,  
Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C.,  
Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,  
Meidrim, J., Menus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J.,  
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,  
Rachupka, A., Ramaamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,  
Roman, J., Schauer, S., Schupack, R., Seaman, S., Severy, P., Smith, C.,  
Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,  
Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M.,  
Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,  
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

Submitted (24-FEB-2003) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 214192)

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,  
Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,  
Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y.,  
Collamore, A., Cook, A., Cooke, P., Corum, B., DeArellano, K.,  
Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,  
Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,  
Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,  
Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,  
Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C.,  
Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,  
Meidrim, J., Menus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J.,  
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,  
Rachupka, A., Ramaamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,  
Roman, J., Schauer, S., Schupack, R., Seaman, S., Severy, P., Smith, C.,  
Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,  
Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M.,  
Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,  
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission

## TITLE

## JOURNAL

## COMMENT

Submitted (14-JUN-2003) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jun 14, 2003 this sequence version replaced gi:30984727.  
All repeats were identified using RepeatMasker:  
Smit, A.P.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L29105

Center clone name: 212\_K\_2

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 211490 bases at least Q40  
Consensus quality: 212337 bases at least Q30  
Consensus quality: 212887 bases at least Q20  
Insert size: 215000; agarose-fp  
Insert size: 213392; sum-of-contigs  
Quality coverage: 10.5 in Q20 bases; agarose-fp  
Quality coverage: 10.6 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 9 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 26288: contig of 26288 bp in length  
\* 26289 26388: gap of 100 bp  
\* 26389 27183: contig of 795 bp in length  
\* 27184 27283: gap of 100 bp  
\* 27284 27997: contig of 713 bp in length  
\* 27997 28096: gap of 100 bp  
\* 28097 29131: contig of 1035 bp in length  
\* 29132 29232: gap of 100 bp  
\* 29232 30463: contig of 1231 bp in length  
\* 30463 33392: contig of 2830 bp in length  
\* 33392 33492: gap of 100 bp  
\* 33493 146267: contig of 112775 bp in length  
\* 146268 146367: gap of 100 bp  
\* 146368 206059: contig of 59692 bp in length  
\* 206060 206159: gap of 100 bp  
\* 206160 214192: contig of 8033 bp in length.

## FEATURES

## source

1. 214192  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="10"  
/map="10"

/clone="RP24-212K2"

/clone\_lib="RPC1-24 Male Mouse BAC"

## misc\_feature

1. 26288  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:left

26289. 26388

/estimated\_length=100

## misc\_feature

26389. 27183

/note="assembly\_fragment"

27184. 27283

/estimated\_length=100

## misc\_feature

27284. 27996

/note="assembly\_fragment"

27997. 28096

/estimated\_length=100

## misc\_feature

28097. 29131

/note="assembly\_fragment"

29132. 29231

/estimated\_length=100

## misc\_feature

29232. 30462

/note="assembly\_fragment"

30463. 30562

/estimated\_length=100

## gap

```

misc_feature 30563..33392
              /note="assembly_fragment"
gap          33393..33492
              /estimated_length=100
misc_feature 33493..146267
              /note="assembly_fragment"
gap          146268..146367
              /estimated_length=100
misc_feature 146368..206059
              /note="assembly_fragment"
gap          206060..206159
              /estimated_length=100
misc_feature 206160..214192
              /note="assembly_fragment"
              clone_end:T7
              vector_side:right"

ORIGIN
Query Match      83.6%; Score 18.4; DB 14; Length 214192;
Best Local Similarity 95.0%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCATGCTCACACATTAACTG 20
   |||||
Db 8528 TCATGCTCACAGATTAACTG 8547

RESULT 23
AC153867
LOCUS
DEFINITION AC153867 196838 bp DNA linear ROD 21-JAN-2005
            Mus musculus 6 BAC RP23-400K22 (Roswell Park Cancer Institute
            (CS7BL/6J Female) Mouse6 BAC Library) complete sequence.

ACCESSION
VERSION AC153867.3 GI:58000476
KEYWORDS
SOURCE Mus musculus (house mouse)

ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidae; Muridae; Murinae; Mus.

Muzny,D., Adams,C., Agbai II,O., Allen,C., Alsbrooks,S., Archer,P.,
Arredondo,H., Bandaranaike,D., Bangura,L., Beltran,B., Beltran,R.,
Beraducci,A., Biswal,K., Blyth,P., Bonham,H., Buhay,C., Burch,P.,
Cadoree,I., Canada,A., Cardenas,V., Carter,K., Cavazos,I.,
Chacko,J., Chahrour,M., Chavez,D., Chen,A., Chen,G., Chen,R.,
Cheng,M.-T., Chu,J., Clerc,K., Cockrell,R., Coyle,M., Cree,A.,
Curry,S., Dai,W., Davila,M.L., Davis,C., Davy-Carroll,L., De
Anda,C., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H.,
Donlin,J., McCauley,S., Dugan-Rocha,S., Dunn,A., Durbin,K.,
Dziuda,D., Egan,A., Escotto,M., Espinosa,V., Eugene,C., Fa,M.,
Fernandez,S., Fernando,P., Flagg,N., Forbes,L., Foster,P.,
Fowler,G., Fu,Q., Fuh,E., Garcia,A., Garcia,R., Garner,T.,
Gaskin,C., Gensch,S., Ghose,S., Gill,R., Gonzalez,D.,
Gonzalez-Garay,M., Guevara,W., Holder,M., Haaland,W., Haeblerlein,K.,
Hall,B., Hamid,H., Hamilton,K., Harbes,B., Harris,R., Havlak,P.,
Hawes,A., Hawkins,E., Hayes,S., Hemphill,L., Hernandez,J.,
Hines,S., Hitchens,M., Hodgson,A., Hogues,M., Hollins,B.,
Howell,L.T., Hulyk,S., Hume,J., Imo,K., Jackson,A., Jackson,L.,
Jacob,L., Jiang,H., Johnson,B., Johnson,R., Kalafas,K., Kelly,S.,
Keys,T., Khan,Z., King,L., Kovar,C., Kowis,A., Kowis,C., Lara,F.,
Leal,S., Lee,K., Lee,S., LeGall,F.I., Lemon,S., Lewis,L., Li,B.,
Li,Y., Li,Z., Linnell,M., Liu,W., Liu,Y.-S., Liu,Y., Liyanage,D.,
London,P., Lopez,J., Lorensuewa,L., Lozados,R., Luk,T., Madu,R.,
Maheshwari,M., Mahoney,C., Malloy,K., Mansouri,D., Martinez,E.,
McCllelland,H., McPherson,J., Mercadao,C., Metzker,M.,
Miloavljivic,A., Minja,E., Morgan,M., Morris,S., Munidasa,M.,
Murray,D., Nazarith,L., Ngo,D., Nguyen,N., Norwig-Eastaugh,E.,
Nott,A., Nwaokelemeh,O., Oregon,M., Ochi-Okorie,C., Oden,E.,
Okwuonu,G., Okwuonu,K., Parker,D., Pasternak,S., Patel,B.,
Patel,V., Paul,H., Perez,A., Perez,L., Petrosino,J., Pham,T.,
Primus,E., Pu,L.-L., Puazo,M., Qin,X., Quinn,A., Quiroz,J.,
Rabata,D., Rachlin,E., Reigh,R., Ren,Y., Reuter,M., Richards,S.,

```

Rives,C., Rodriguez,F., Rojas,A., Ruiz,S.J., Sana,M., Sanders,W.,  
Santibanez,J., Santos,R., Savery,G., Scherer,S., Shen,H., Shen,Y.,  
Sisson,I., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R.,  
Svatek,A., Taylor,E., Taylor,T., Thomas,N., Thorn,R., Thornton,R.,  
Trejos,Z., Umani,K., Vargo,C., Verduzco,D., Villasana,D., Virk,D.,  
Volkov,A., Waldron,L., Walker,B., Wang,Q., Wang,S., Warren,J.,  
Wei,X., Wheeler,D., Williams,G., Williams,R., Worley,K., Wright,R.,  
Wu,J., Yakub,S., Yan,K., Yaun,Y., Yu,F., Zhang,J., Zhang,L.,  
Zhang,Z., Zhou,J., Weinstock,G. and Gibbs,R.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Direct Submission  
Unpublished  
2 (bases 1 to 196838)  
Worley,K.C.  
Direct Submission  
Submitted (18-DEC-2004) Human Genome Sequencing Center, Baylor  
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 196838)  
Worley,K.C.  
Direct Submission  
Submitted (13-JAN-2005) Human Genome Sequencing Center, Baylor  
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
4 (bases 1 to 196838)  
Worley,K.C.  
Direct Submission  
Submitted (21-JAN-2005) Human Genome Sequencing Center, Baylor  
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Jan 21, 2005 this sequence version replaced gi:57634374.  
Sequencing is completed to a minimum standard of double strand  
coverage with a minimum of 2 clones and 2 reads with no ambiguities  
or 2 chemistries with a minimum of 2 clones and 3 reads with no  
ambiguities. If the sequence quality does not meet this standard,  
it will be indicated in the annotation.

The repeat regions shown were identified using RepeatMasker by  
Adrian Smit.

Sequence similarities were identified using Powerblast by Jinghui  
Zhang.

Exon/Intron boundaries of identified genes were chosen if there  
were canonical splice junctions that maintained sequence continuity  
across the splice junctions.

FEATURES  
Source  
Location/Qualifiers  
1..196838  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/clone="RP23-400K22"  
1..98  
/rpt\_family="B4"  
666..841  
/rpt\_family="MER2"  
complement(1719..1939)  
/rpt\_family="URR1B"  
1940..1979  
/rpt\_family="CYRA11.Mm"  
complement(2988..3132)  
/rpt\_family="B4"  
complement(3134..3255)  
/rpt\_family="Lx7"  
3256..3322  
/rpt\_family="(TC)n"  
complement(3323..3546)  
/rpt\_family="Lx7"  
3552..3626  
/rpt\_family="GA-rich"  
6140..6248  
/rpt\_family="PB1D10"  
6580..6703  
/rpt\_family="MIR"  
7344..7450  
/rpt\_family="L2"  
8694..8860

```
repeat_region /rpt_family="L1ME1"
complement(8930..9068)
repeat_region /rpt_family="ID_B1"
9081..9199
repeat_region /rpt_family="L1ME1"
9201..9473
repeat_region /rpt_family="L1ME1"
12449..12493
repeat_region /rpt_family="(TC)n"
complement(113843..113947)
repeat_region /rpt_family="MER112"
14436..14700
repeat_region /rpt_family="RMR20B"
15772..15985
repeat_region /rpt_family="B4"
complement(15986..16369)
repeat_region /rpt_family="MTC"
16370..16421
repeat_region /rpt_family="B4"
16463..16561
repeat_region /rpt_family="(CA)n"
complement(16727..17532)
repeat_region /rpt_family="Charlie7"
17555..17585
repeat_region /rpt_family="AT_rich"
17903..17994
repeat_region /rpt_family="Lx9"
18108..18246
repeat_region /rpt_family="L1_Mus4"
18247..18422
repeat_region /rpt_family="B3"
19813..19835
repeat_region /rpt_family="AT_rich"
20561..20775
repeat_region /rpt_family="MLT1A1"
20860..20963
repeat_region /rpt_family="(TTC)n"
complement(21016..21130)
repeat_region /rpt_family="B2_Mm2"
21284..21364
repeat_region /rpt_family="MTE"
21656..22133
repeat_region /rpt_family="MT2A"
22346..22672
repeat_region /rpt_family="ORR1C2"
24405..24448
repeat_region /rpt_family="(CAAAA)n"
24669..24703
repeat_region /rpt_family="(TTTA)n"
24705..24727
repeat_region complement(24763..24973)
/rpt_family="B3"
26848..27365
repeat_region /rpt_family="RLTR11B"
28774..28817
repeat_region /rpt_family="(TC)n"
complement(29086..29210)
repeat_region /rpt_family="L2"
complement(29346..29636)
repeat_region /rpt_family="L2"
29797..30221
repeat_region /rpt_family="MTD"
complement(30366..30719)
/rpt_family="L2"
31112..31240
repeat_region /rpt_family="MER58B"
31475..31816
repeat_region /rpt_family="MTD"
complement(31817..32004)
repeat_region /rpt_family="B2_Mm1"
32005..32037
repeat_region /rpt_family="MTD"
```

```
repeat_region 32282..32469
/rpt_family="MTE"
repeat_region 32535..32641
/rpt_family="MTE"
repeat_region complement(34702..35054)
/rpt_family="ORR1D"
repeat_region 36300..36363
/rpt_family="(TCTA)n"
complement(37295..37402)
/rpt_family="MIR"

Query Match 81.8%; Score 18; DB 9; Length 196838;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCATGCTCACATTAAAC 18
Db 170720 TCATGCTCACATTAAAC 170737

RESULT 24
AC153630
LOCUS AC153630
DEFINITION Mus musculus 6 BAC RP23-289016 (Roswell Park Cancer Institute
(C57BL/6J Female) Mouse BAC Library) complete sequence.
ACCESSION AC153630 AC102141
VERSION AC153630.1 GI:56565712
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
AUTHORS Muzny,D., Adams,C., Agbai II,O., Allen,C., Alsbrooks,S., Archer,P.,
Arredondo,H., Bandaranaike,D., Bangura,L., Beltran,B., Beltran,R.,
Beraducci,A., Biewal,K., Blyth,P., Bonham,H., Buhay,C., Burch,P.,
Cadoree,I., Canada,A., Cardenas,V., Carter,K., Cavazos,I.,
Chacko,J., Chahrouh,M., Chavez,D., Chen,A., Chen,G., Chen,R.,
Cheng,M.-T., Chu,J., Clerc,K., Cockrell,R., Coyle,M., Cree,A.,
Curry,S., Dai,W., Davila,M.B., Davis,C., Davy-Carroll,L., De
Anda,C., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H.,
Doblin,J., McCauley,S., Dugan-Rocha,S., Dunn,A., Durbin,K.,
Dziuda,D., Egan,A., Escotto,M., Espinosa,V., Eugene,C., Fa,M.,
Fernandez,S., Fernando,P., Flagg,N., Forbes,L., Foster,P.,
Fowler,G., Fu,Q., Fuh,E., Garcia,A., Garcia,R., Garner,T.,
Gaskin,C., Gench,S., Ghose,S., Gill,R., Gonzalez,D.,
Gonzalez-Garay,M., Guevara,W., Holder,M., Haaland,W., Haerlein,K.,
Hall,B., Hamid,H., Hamilton,K., Harbes,B., Harris,R., Havlak,P.,
Hawes,A., Hawkins,E., Hayes,S., Hemphill,L., Hernandez,J.,
Hines,S., Hitchens,M., Hodgeson,A., Hughes,M., Hollins,B.,
Howell,L.T., Hulyk,S., Hume,J., Imo,K., Jackson,A., Jackson,L.,
Jacobs,L., Jiang,H., Johnson,B., Johnson,R., Kalafus,K., Kelly,S.,
Keys,T., Khan,Z., King,L., Kovar,C., Kowis,A., Kowis,C., Lara,F.,
Leal,S., Lee,K., Lee,S., LeGall,F.I., Lemon,S., Lewis,L., Li,B.,
Li,Y., Li,Z., Linnell,M., Liu,W., Liu,Y.-S., Liu,Y., Liyanage,D.,
London,P., Lopez,J., Lorensuhewa,L., Lozada,R., Luk,T., Madu,R.,
Maheshwari,M., Mahoney,C., Malloy,K., Mansouri,D., Martinez,E.,
McClelland,H., McPherson,J., Mercadao,C., Metzker,M.,
Milosavljevic,A., Minja,E., Morgan,M., Morris,S., Muntadasa,M.,
Murray,D., Nazareth,L., Ngo,D., Nguyen,N., Norwig-Eastaugh,E.,
Nott,A., Nwaokemele,O., Oregon,M., Ochi-Okorie,C., Odeh,E.,
Okwuonu,G., Okwuonu,K., Parker,D., Pasternak,S., Patel,B.,
Patel,V., Paul,H., Perez,A., Perez,L., Petrosino,J., Pham,T.,
Primus,E., Pu,L.-L., Puazo,M., Qin,X., Quinn,A., Quiros,J.,
Rabata,D., Rachlin,E., Reigh,R., Ren,Y., Reuter,M., Richards,S.,
Rives,C., Rodriguez,F., Rojas,A., Ruiz,S.J., Sana,M., Sanders,W.,
Santibanez,J., Santos,R., Savery,G., Scherer,S., Shen,H., Shen,Y.,
Sisson,I., Sneed,A., Sodergren,E., Song,X.-Z., Sorrelle,R.,
Svatek,A., Taylor,E., Taylor,T., Thomas,N., Thorn,R., Thornton,R.,
Trejos,Z., Usmani,K., Vargo,C., Verduzco,D., Villana,D., Virk,D.,
Volkov,A., Waldron,L., Walker,B., Wang,Q., Wang,S., Warren,J.,
Wei,X., Wheeler,D., Williams,G., Williams,R., Worley,K., Wright,R.,
```





```

repeat_region 42674..42878
/rpt_family="ID_B1"
repeat_region 42879..43257

Query Match      81.8%  Score 18;  DB 9;  Length 217496;
Best Local Similarity 100.0%;  Pred. No. 2.4e+02;
Matches 18;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

OY 1 TCATGCTCACACATTAAAC 18
|||||
Db 117030 TCATGCTCACACATTAAAC 117047

RESULT 25
AC158678/c
LOCUS
DEFINITION
AC158678 217586 bp DNA linear ROD 30-MAY-2005
Mus musculus 6 BAC RP23-14216 (Roswell Park Cancer Institute
(C57BL/6J Female) Mouse BAC Library) complete sequence.
AC158678
HTG.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 217586)
Muzny, D., Adams, C., Agbai II, O., Allen, C., Alsbrooks, S., Archer, P.,
Arredondo, H., Bandaranaike, D., Bangura, L., Beltran, B., Beltran, R.,
Berarducci, A., Biswal, K., Blyth, P., Bonham, H., Buhay, C., Burch, P.,
Cadoree, J., Canada, A., Cardenas, V., Carter, K., Cavazos, I.,
Chacko, J., Chahrour, M., Chavez, D., Chen, A., Chen, G., Chen, R.,
Cheng, M.-T., Chu, J., Clerc, K., Cockrell, R., Coyle, M., Cree, A.,
Curry, S., Dai, W., Davila, M.L., Davis, C., Davy-Carroll, L., De
Anda, C., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H.,
Donlin, J., McCauley, S., Dugan-Rocha, S., Dunn, A., Durbin, K.,
Dzulida, D., Egan, A., Escotto, M., Espinosa, V., Eugene, C., Fa, M.,
Fernandez, S., Fernando, P., Flagg, N., Forbes, L., Foster, P.,
Fowler, G., Fu, Q., Fuh, E., Garcia, A., Garcia, R., Garner, T.,
Gaskin, C., Gench, S., Ghose, S., Gill, R., Gonzalez, D.,
Gonzalez-Garay, M., Guevara, W., Holder, M., Hasland, W., Haeblerlein, K.,
Hall, B., Hamid, H., Hamilton, K., Harbes, B., Harris, R., Havlak, P.,
Hawes, A., Hawkins, E., Hayes, S., Hodgson, A., Hogue, M., Hollins, B.,
Hines, S., Hitchens, M., Hodgson, A., Hogue, M., Hollins, B.,
Howell, L.T., Hulyk, S., Hume, J., Imo, K., Jackson, A., Jackson, L.,
Jacob, L., Jiang, H., Johnson, B., Johnson, R., Kalafus, K., Kelly, S.,
Keys, T., Khan, Z., King, L., Kovar, C., Kowis, A., Kowis, C., Lara, P.,
Leal, S., Lee, K., Lee, S., Legall, F.I., Lemon, S., Lewis, L., Li, B.,
Li, Y., Li, Z., Linnell, M., Liu, W., Liu, Y.-S., Liu, Y., Liyanage, D.,
London, P., Lopez, J., Lorenshewa, L., Lozado, R., Luk, T., Madu, R.,
Maheshwari, M., Mahoney, C., Malloy, K., Mansouri, D., Martinez, E.,
McClelland, H., McPherson, J., Mercadao, C., Metzker, M.,
Milosavljevic, A., Minja, E., Morgan, M., Morris, S., Munidasa, M.,
Murray, D., Nazareth, L., Ngo, D., Nguyen, N., Norwig-Eastaugh, E.,
Nott, A., Nwaokemele, O., Obregon, M., Ochi-Okorie, C., Odeh, E.,
Okwuonu, G., Okwuonu, K., Parker, D., Pasternak, S., Patel, B.,
Patel, V., Paul, H., Perez, A., Perez, L., Petrosino, J., Pham, T.,
Primus, E., Pu, L.-L., Puzos, M., Qin, X., Quinn, A., Quiroz, J.,
Rabata, D., Rachlin, E., Reign, R., Ren, Y., Reuter, M., Richards, S.,
Rives, C., Rodriguez, F., Rojas, A., Ruiz, S.-J., Sana, M., Sanders, W.,
Santibanez, J., Santos, R., Savary, G., Scherer, S., Shen, H., Shen, Y.,
Sisson, I., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R.,
Svatek, A., Taylor, E., Taylor, T., Thomas, N., Thorn, R., Thornton, R.,
Trejos, Z., Usmani, K., Vargo, C., Verduzco, D., Villasana, D., Virk, D.,
Voikov, A., Waldron, L., Walker, B., Wang, Q., Wang, S., Warren, J.,
Wei, X., Wheeler, D., Williams, G., Williams, R., Worley, K., Wright, R.,
Wu, J., Yakub, S., Yan, K., Yuan, Y., Yu, F., Zhang, J., Zhang, L.,
Zhang, Z., Zhou, J., Weinstock, G. and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 217586)
Worley, K.C.
Direct Submission
JOURNAL
REFERENCE
AUTHORS
TITLE

```

Submitted (23-MAR-2005) Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 217586)  
Worley, K.C.  
Direct Submission  
Submitted (07-MAY-2005) Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
4 (bases 1 to 217586)  
Worley, K.C.  
Direct Submission  
Submitted (30-MAY-2005) Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On May 30, 2005 this sequence version replaced gi:53094003.  
Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality does not meet this standard, it will be indicated in the annotation.

The repeat regions shown were identified using RepeatMasker by Adrian Smit.

Sequence similarities were identified using PowerBlast by Jinghui Zhang.

Exon/Intron boundaries of identified genes were chosen if there were canonical splice junctions that maintained sequence continuity across the splice junctions.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Drafting Center Code: BCM  
Contact: hgsc-help@bcm.tmc.edu.  
Location/Qualifiers  
1..217586  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosomes="6"  
/clone="RP23-14216"  
complement(1..75150)  
/note="overlaps bases 1..75142 of clone AC153841"  
-----  
1..43  
/rpt\_family="ID\_B1"  
complement(1169..1577)  
/rpt\_family="MLT1C"  
complement(1662..7937)  
/rpt\_family="LIMD\_F2"  
complement(8396..8587)  
/rpt\_family="B3A"  
8659..8882  
/rpt\_family="MIR"  
10336..10604  
/rpt\_family="B4A"  
11621..11664  
/rpt\_family="CAAAA)n"  
11726..11790  
/rpt\_family="(TG)n"  
complement(11809..11975)  
/rpt\_family="RSINEL"  
11981..12002  
/rpt\_family="AT rich"  
complement(12009..12085)  
/rpt\_family="ID4"  
12181..12205  
/rpt\_family="(TTG)n"  
12508..12903  
/rpt\_family="MTA\_Mm"  
14033..14515  
/rpt\_family="RLTR16"  
15256..15294

/rpt family="AT rich"  
complement(15299..16229)  
/rpt family="Li\_Musi"  
16315..16599  
/rpt family="LiM4"  
17168..17202  
/rpt family="AT rich"  
18762..18861  
/rpt family="GA-rich"  
20374..20499  
/rpt family="BIF"  
21419..21499  
/rpt family="TCCC)n"  
21461..21465  
/function="single clone coverage"  
21508..21573  
/rpt family="TCCA)n"  
21772..21814  
/rpt family="AT rich"  
complement(22758..24332)  
/rpt family="Li\_Musi"  
complement(24554..24900)  
/rpt family="MLT1B"  
27571..27652  
/rpt family="TATATG)n"  
28033..28323  
/rpt family="MLT1J"  
28578..28653  
/rpt family="MTE"  
28733..28887  
/rpt family="MTE"  
28895..29556  
/rpt family="MTE-int"  
29631..30326  
/rpt family="MTE-int"  
30329..30731  
/rpt family="MTE"  
30989..31299  
/rpt family="L2"  
31966..32049  
/rpt family="CT-rich"  
complement(32735..33101)  
/rpt family="MTE"  
complement(33534..34031)  
/rpt family="L3"  
34384..34590  
/rpt family="ORR1A3"  
34591..34912  
/rpt family="RLTRETN\_Mm"  
34913..35036  
/rpt family="ORR1A3"  
complement(35346..35420)  
/rpt family="L3"  
complement(35703..35914)  
/rpt family="B3"  
36029..36116  
/rpt family="MLT1A1"  
36117..36186  
/rpt family="TGT)n"  
36187..36425  
/rpt family="MLT1A1"  
complement(36541..36639)  
/rpt family="L3"  
36664..36725  
/rpt family="(TTA)n"  
37558..37909  
/rpt family="ORR1D"  
38474..38584  
/rpt family="MLT1A1"

Query Match 81.8%; Score 18; DB 9; Length 217586;  
Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCATGCTCACATTAAAC 18  
|||||  
Db 186204 TCATGCTCACATTAAAC 186187  
|||||  
RESULT 26  
AB014686 4363 bp DNA linear BCT 23-JUL-2003  
LOCUS Streptococcus bovis gene for pyruvate formate-lyase, complete cds.  
DEFINITION AB014686  
ACCESSION AB014686  
VERSION AB014686.1 GI:3168595  
KEYWORDS pyruvate formate-lyase.  
SOURCE Streptococcus bovis  
ORGANISM Streptococcus bovis  
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
Streptococcus.  
1  
REFERENCE  
AUTHORS Asanuma,N., Iwamoto,M. and Hino,T.  
TITLE Structure and transcriptional regulation of the gene encoding  
pyruvate formate-lyase of a ruminal bacterium, Streptococcus bovis  
JOURNAL Microbiology (Reading, Engl.) 145 (Pt 1), 151-157 (1999)  
PUBMED 10206694  
REFERENCE 2 (bases 1 to 4363)  
AUTHORS Asanuma,N., Iwamoto,M. and Hino,T.  
TITLE Direct Submission  
JOURNAL Submitted (24-MAY-1998) Tsunao Hino, Meiji University, Department  
of Agriculture; Higashimita Tama-Ku, Kawasaki, Kanagawa 214-8571,  
Japan (E-mail:asanuma@isc.meiji.ac.jp, Tel: +81-44-934-7895,  
Fax: +81-44-934-7825)  
FEATURES  
Location/Qualifiers  
1..4363  
/organism="Streptococcus bovis"  
/mol\_type="Genomic DNA"  
/strain="JB-1"  
/db\_xref="taxon:1315"  
1707..4031  
/gene="pfl"  
1707..4031  
/gene="pfl"  
/codon\_start=1  
/transl\_table=11  
/product="pyruvate formate-lyase"  
/protein\_id="BAA28614.1"  
/db\_xref="GI:3168596"  
/translation="MATVKTADVFEKAWGFKGTDWKEKASVSRFQVANYTPYDGD  
SFLAPATERSLKIKKIIEDTKAEYEAATPFMDTRPTSADIAPAGIQKDDLELYIQN  
DELFLNMPKGGIRMAETALKEHGYEPDPAVHEIFTKYTTTNDGI FRAVTSNTRRA  
RHAHTVTGLPDAYSRRGIIGVVARLALYADYLMQEVNDWNAITEIDEEISRLREEV  
NMVOALGEVVKGLDLYGVKPPAMNVKEAIQWNIAMVAVCRVINGAATSLGRVPI  
VLDDI PAERDLARGTTESEIQEFVDDFVLKLTVPKARTKAYDELYSGDPTTITISMA  
GMGADRHRTKMDYRFLHTLDNIGNAPEPLNLTWTKLPYSFRYCKMMSHKISSI  
QYEGVTTMAKDGYGEMSCISCCVPLDPENEEQRHNI QYFGARVNLKALLTGLNGGY  
DDVHKDYFOIPEYVRDILDPETVKANFEKSLDWLTSTYVDALNI IHYMTDKYNEA  
VQMAFLPTKORANMGFGICGPANTVDTLSAIKYATVKPIRDEDDGIYDYETTGDFPRW  
GEDPERSNELAEWLVEATTRLRSHKLYKNAEATVSLLTITSNVAYSKOTGNSPVHKG  
VYINEDSGTNLSKLEFFPFGANPSKARGCNLQNLNSLASLDFSAADGISITTTQVSP  
RALGKTDFEQDNLVTILDGYPENGQHVNLNMDLKQVIDKIMMGEDVIVRISGICV  
NTKYLTKEQKTELTQRFHEVLSDMDVAFTAAK"

Query Match 80.9%; Score 17.8; DB 1; Length 4363;  
Best Local Similarity 90.5%; Pred. No. 4.2e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCATGCTCACATTAACTGG 21  
|||||  
Db 2189 TCATGCTCACATTAACTGG 2209  
|||||

RESULT 27  
AC010512/c  
LOCUS AC010512 AC010512 39915 bp DNA linear PRI 13-JUL-2002

```

DEFINITION Homo sapiens chromosome 19 clone LLNLR-226F1, complete sequence.
ACCESSION AC010512
VERSION AC010512.7 GI:8099261
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homnidae; Homo.
REFERENCE 1 (bases 1 to 39915)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 39915)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 39915)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (27-MAY-2000) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
REFERENCE 4 (bases 1 to 39915)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (13-JUL-2002) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On May 27, 2000 this sequence version replaced gi:7711461.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99% of Sequence;
Estimated Total Number of Errors is 0.5.
FEATURES             source
     1..39915
        Location/Qualifiers
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="19"
            /clone="LLNLR-226F1"
ORIGIN
Query Match      80.9%; Score 17.8; DB 8; Length 39915;
Best Local Similarity 90.5%; Pred. No. 3.5e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCATGCTCACATTAACTGG 21
    ||||| ||||| ||||| |||||
Db 10108 TCATGCTCACATTGACTGG 10088

RESULT 28
AC021693/c
LOCUS AC021693 Homo sapiens clone RP11-23J10, LOW-PASS SEQUENCE SAMPLING.
DEFINITION AC021693
ACCESSION AC021693
VERSION AC021693.2 GI:9128545
KEYWORDS HTG; HTGS_PHASE0.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homnidae; Homo.
REFERENCE 1 (bases 1 to 72147)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-23J10
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 72147)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavskiy,L., Boukhalter,B., Brown,A., Burkett,G., Castle,A.,

```

# TITLE JOURNAL COMMENT

Choepe,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,  
DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,  
Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,  
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
Landers,T., Leloczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,  
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,  
McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,  
Norman,C.H., O'Connor,T., O'Donnell,P., Olivari,T.M., Peterson,K.,  
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,  
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,  
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,  
Zimmer,A. and Zody,M.

Direct Submission  
Submitted (19-JAN-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jul 13, 2000 this sequence version replaced gi:6715892.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L4378  
Center clone name: 23\_J\_10  
-----

\* NOTE: This record contains 87 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 704: contig of 704 bp in length  
\* 705 804: gap of 100 bp  
\* 805 1520: contig of 716 bp in length  
\* 1521 1620: gap of 100 bp  
\* 1621 2346: contig of 726 bp in length  
\* 2347 2446: gap of 100 bp  
\* 2447 3190: contig of 744 bp in length  
\* 3191 3290: gap of 100 bp  
\* 3291 4025: contig of 735 bp in length  
\* 4026 4125: gap of 100 bp  
\* 4126 4845: contig of 720 bp in length  
\* 4846 4945: gap of 100 bp  
\* 4946 5676: contig of 731 bp in length  
\* 5677 5776: gap of 100 bp  
\* 5777 6507: contig of 731 bp in length  
\* 6508 6607: gap of 100 bp  
\* 6608 7331: contig of 724 bp in length  
\* 7332 7432: gap of 100 bp  
\* 7432 8156: contig of 725 bp in length  
\* 8157 8256: gap of 100 bp  
\* 8257 8968: contig of 712 bp in length  
\* 8969 9068: gap of 100 bp  
\* 9069 9809: contig of 741 bp in length  
\* 9810 9909: gap of 100 bp  
\* 9910 10642: contig of 733 bp in length  
\* 10643 10742: gap of 100 bp  
\* 10743 11465: contig of 723 bp in length  
\* 11466 11565: gap of 100 bp  
\* 11566 12307: contig of 742 bp in length  
\* 12308 12407: gap of 100 bp  
\* 12408 13140: contig of 733 bp in length  
\* 13141 13240: gap of 100 bp

```
* 13241 13978: contig of 738 bp in length
* 13979 14078: gap of 100 bp
* 14079 14804: contig of 726 bp in length
* 14805 14904: gap of 100 bp
* 14905 15625: contig of 721 bp in length
* 15626 15725: gap of 100 bp
* 15726 16454: contig of 729 bp in length
* 16455 16554: gap of 100 bp
* 16555 17281: contig of 727 bp in length
* 17282 17381: gap of 100 bp
* 17382 18123: contig of 742 bp in length
* 18124 18223: gap of 100 bp
* 18224 18949: contig of 726 bp in length
* 18950 19049: gap of 100 bp
* 19050 19780: contig of 731 bp in length
* 19781 19880: gap of 100 bp
* 19881 20623: contig of 743 bp in length
* 20624 20723: gap of 100 bp
* 20724 21457: contig of 734 bp in length
* 21458 21557: gap of 100 bp
* 21558 22236: contig of 739 bp in length
* 22237 22396: gap of 100 bp
* 22397 23126: contig of 730 bp in length
* 23127 23226: gap of 100 bp
* 23227 23953: contig of 727 bp in length
* 23954 24053: gap of 100 bp
* 24054 24774: contig of 721 bp in length
* 24775 24874: gap of 100 bp
* 24875 25612: contig of 738 bp in length
* 25613 25712: gap of 100 bp
* 25713 26445: contig of 733 bp in length
* 26446 26545: gap of 100 bp
* 26546 27289: contig of 724 bp in length
* 27270 27369: gap of 100 bp
* 27370 28119: contig of 750 bp in length
* 28120 28219: gap of 100 bp
* 28220 28954: contig of 735 bp in length
* 28955 29054: gap of 100 bp
* 29055 29792: contig of 738 bp in length
* 29793 29892: gap of 100 bp
* 29893 30629: contig of 737 bp in length
* 30630 30729: gap of 100 bp
* 30730 31445: contig of 716 bp in length
* 31446 32289: gap of 100 bp
* 31546 32289: contig of 724 bp in length
* 32270 32369: gap of 100 bp
* 32370 33099: contig of 730 bp in length
* 33100 33199: gap of 100 bp
* 33200 33933: contig of 734 bp in length
* 33934 34033: gap of 100 bp
* 34034 34781: contig of 728 bp in length
* 34762 34861: gap of 100 bp
* 34862 35574: contig of 713 bp in length
* 35575 35674: gap of 100 bp
* 35675 36410: contig of 736 bp in length
* 36411 36510: gap of 100 bp
* 36511 37230: contig of 720 bp in length
* 37231 37330: gap of 100 bp
* 37331 38053: contig of 723 bp in length
* 38054 38153: gap of 100 bp
* 38154 38886: contig of 733 bp in length
* 38887 38986: gap of 100 bp
* 38987 39741: contig of 755 bp in length
* 39742 39841: gap of 100 bp
* 39842 40553: contig of 712 bp in length
* 40554 40654: gap of 100 bp
* 40654 41389: contig of 735 bp in length
* 41389 41488: gap of 100 bp
* 41489 42220: contig of 732 bp in length
* 42221 42320: gap of 100 bp
* 42321 43084: contig of 764 bp in length
* 43085 43184: gap of 100 bp
* 43185 43914: contig of 730 bp in length
```

```
* 43915 44014: gap of 100 bp
* 44015 44763: contig of 749 bp in length
* 44764 44863: gap of 100 bp
* 44864 45587: contig of 724 bp in length
* 45588 45687: gap of 100 bp
* 45688 46424: contig of 737 bp in length
* 46425 46524: gap of 100 bp
* 46525 47260: contig of 736 bp in length
* 47261 47360: gap of 100 bp
* 47361 48100: contig of 740 bp in length
* 48101 48200: gap of 100 bp
* 48201 48928: contig of 728 bp in length
* 48929 49028: gap of 100 bp
* 49029 49754: contig of 726 bp in length
* 49755 49854: gap of 100 bp
* 49855 50583: contig of 729 bp in length
* 50584 50683: gap of 100 bp
* 50684 51415: contig of 732 bp in length
* 51416 51515: gap of 100 bp
* 51516 52245: contig of 730 bp in length
* 52246 52345: gap of 100 bp
* 52346 53051: contig of 706 bp in length
* 53052 53151: gap of 100 bp
* 53152 53877: contig of 726 bp in length
* 53878 53977: gap of 100 bp
* 53978 54695: contig of 718 bp in length
* 54696 54795: gap of 100 bp
* 54796 55525: contig of 730 bp in length
* 55526 55625: gap of 100 bp
* 55626 56365: contig of 740 bp in length
* 56366 56465: gap of 100 bp
* 56466 57205: contig of 740 bp in length
* 57206 57305: gap of 100 bp
* 57306 58037: contig of 732 bp in length
* 58038 58137: gap of 100 bp
* 58138 58865: contig of 728 bp in length
```

Query Match 80.9%; Score 17.8; DB 14; Length 72147;

Best Local Similarity 90.5%; Pred. No. 3 3e+02;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy      2 CATGCTCACACATTAACCTGA 22
          ||||| ||||| ||||| |||||
Db      26011 CATGCTCAATCAATTAACCTGA 25991
```

#### RESULT 29

```
AC016504
LOCUS      Homo sapiens clone RP11-7L18, LOW-PASS SEQUENCE SAMPLING.
DEFINITION AC016504.2 GI:91233447
ACCESSION  AC016504
VERSION     HTG; HTGS PHASE0.
KEYWORDS    Homo sapiens (human)
SOURCE      Homo sapiens
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homnidae; Homo.
REFERENCE   1 (bases 1 to 87619)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens, clone RP11-7L18
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 87619)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
            Brown,A., Castle,A., Collings,S., Collamore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Hearford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
```

Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,  
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
 Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,  
 Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

# TITLE JOURNAL

Submitted (01-DEC-1999) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Jul 13, 2000 this sequence version replaced gi:6492523.

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L2949

Center clone name: 7\_L\_18

-----

\* NOTE: This record contains 87 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlapping relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.

\* 1 940: contig of 940 bp in length

\* 941 1040: gap of 100 bp

\* 1041 1958: contig of 918 bp in length

\* 1959 2058: gap of 100 bp

\* 2059 2970: contig of 912 bp in length

\* 2971 3070: gap of 100 bp

\* 3071 3969: contig of 899 bp in length

\* 3970 4069: gap of 100 bp

\* 4070 4977: contig of 908 bp in length

\* 4978 5077: gap of 100 bp

\* 5078 5969: contig of 892 bp in length

\* 5970 6089: gap of 100 bp

\* 6070 6984: contig of 915 bp in length

\* 6985 7084: gap of 100 bp

\* 7085 8000: contig of 916 bp in length

\* 8001 8100: gap of 100 bp

\* 8101 9023: contig of 923 bp in length

\* 9024 9123: gap of 100 bp

\* 9124 10074: contig of 951 bp in length

\* 10075 10174: gap of 100 bp

\* 10175 11079: contig of 905 bp in length

\* 11080 11179: gap of 100 bp

\* 11180 12099: contig of 920 bp in length

\* 12100 12199: gap of 100 bp

\* 12200 13144: contig of 945 bp in length

\* 13145 13244: gap of 100 bp

\* 13245 14166: contig of 922 bp in length

\* 14167 14266: gap of 100 bp

\* 14267 15184: contig of 918 bp in length

\* 15185 15284: gap of 100 bp

\* 15285 16163: contig of 879 bp in length

\* 16164 16263: gap of 100 bp

\* 16264 17184: contig of 921 bp in length

\* 17185 17285: gap of 100 bp

\* 17285 18203: contig of 919 bp in length

\* 18204 18303: gap of 100 bp

\* 18304 19051: contig of 748 bp in length

\* 19052 19151: gap of 100 bp

\* 19152 20051: contig of 900 bp in length

\* 20052 20151: gap of 100 bp

\* 20152 21071: contig of 920 bp in length

\* 21072 21171: gap of 100 bp

\* 21171 21171: gap of 100 bp

\* 21072 21171: gap of 100 bp

\* 21171 21171: gap of 100 bp

\* 21171 21171: gap of 100 bp

\* 21171 21171: gap of 100 bp

\* 21171 21171: gap of 100 bp

\* 21171 21171: gap of 100 bp

\* 21171 21171: gap of 100 bp

22090: contig of 919 bp in length  
 22190: gap of 100 bp  
 23114: contig of 924 bp in length  
 23214: gap of 100 bp  
 24119: contig of 905 bp in length  
 24219: gap of 100 bp  
 25117: contig of 898 bp in length  
 25217: gap of 100 bp  
 26040: contig of 823 bp in length  
 26141: gap of 100 bp  
 27020: contig of 880 bp in length  
 27120: gap of 100 bp  
 28007: contig of 887 bp in length  
 28107: gap of 100 bp  
 28980: contig of 873 bp in length  
 29080: gap of 100 bp  
 29081: 29085: contig of 905 bp in length  
 30085: gap of 100 bp  
 30989: contig of 904 bp in length  
 31089: gap of 100 bp  
 32059: contig of 970 bp in length  
 32159: gap of 100 bp  
 33083: contig of 924 bp in length  
 33183: gap of 100 bp  
 34107: contig of 924 bp in length  
 34207: gap of 100 bp  
 35114: contig of 907 bp in length  
 35214: gap of 100 bp  
 36127: contig of 913 bp in length  
 36227: gap of 100 bp  
 37117: contig of 890 bp in length  
 37217: gap of 100 bp  
 38117: contig of 900 bp in length  
 38217: gap of 100 bp  
 39138: contig of 921 bp in length  
 39238: gap of 100 bp  
 40137: contig of 899 bp in length  
 40237: gap of 100 bp  
 41136: contig of 899 bp in length  
 41236: gap of 100 bp  
 42133: contig of 897 bp in length  
 42233: gap of 100 bp  
 43137: contig of 904 bp in length  
 43237: gap of 100 bp  
 44170: contig of 933 bp in length  
 44270: gap of 100 bp  
 45176: contig of 906 bp in length  
 45276: gap of 100 bp  
 46208: contig of 932 bp in length  
 46308: gap of 100 bp  
 47232: contig of 924 bp in length  
 47332: gap of 100 bp  
 48237: contig of 905 bp in length  
 48337: gap of 100 bp  
 49242: contig of 905 bp in length  
 49342: gap of 100 bp  
 50242: contig of 900 bp in length  
 50342: gap of 100 bp  
 51243: contig of 901 bp in length  
 51343: gap of 100 bp  
 52234: contig of 891 bp in length  
 52334: gap of 100 bp  
 53258: contig of 924 bp in length  
 53358: gap of 100 bp  
 54266: contig of 908 bp in length  
 54366: gap of 100 bp  
 55279: contig of 913 bp in length  
 55379: gap of 100 bp  
 56297: contig of 918 bp in length  
 56397: gap of 100 bp  
 57319: contig of 922 bp in length  
 57419: gap of 100 bp  
 58335: contig of 916 bp in length

\* 58336 58435: gap of 100 bp  
\* 58336 59333: contig of 898 bp in length  
\* 59333 59434: gap of 100 bp  
\* 59434 60356: contig of 923 bp in length  
\* 60356 60456: gap of 100 bp  
\* 60456 61375: contig of 919 bp in length  
\* 61375 61475: gap of 100 bp  
\* 61475 62363: contig of 888 bp in length  
\* 62363 62463: gap of 100 bp  
\* 62463 63380: contig of 917 bp in length  
\* 63380 63480: gap of 100 bp  
\* 63480 64392: contig of 912 bp in length  
\* 64392 64492: gap of 100 bp  
\* 64492 65403: contig of 911 bp in length  
\* 65403 65503: gap of 100 bp  
\* 65503 66430: contig of 927 bp in length  
\* 66430 66530: gap of 100 bp  
\* 66530 67426: contig of 896 bp in length  
\* 67426 67526: gap of 100 bp  
\* 67526 68450: contig of 924 bp in length  
\* 68450 68550: gap of 100 bp  
\* 68550 69457: contig of 907 bp in length  
\* 69457 69557: gap of 100 bp  
\* 69557 70483: contig of 926 bp in length  
\* 70483 70584: gap of 100 bp  
\* 70584 71496: contig of 913 bp in length  
\* 71496 71597: gap of 100 bp  
\* 71597 72525: contig of 929 bp in length  
\* 72525 72625: gap of 100 bp

Query Match 80.9%; Score 17.8; DB 14; Length 87619;  
Best Local Similarity 90.5%; Pred. No. 3.2e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCATGCTCACACATTAACCTGG 21  
||||| ||||| ||||| ||||| |||||  
Db 10278 TCATGACACACAGTAACCTGG 10298

RESULT 30  
AC022755/c  
LOCUS AC022755 92290 bp DNA linear HTG 13-JUL-2000  
DEFINITION Homo sapiens clone RP11-26007, LOW-PASS SEQUENCE SAMPLING.  
ACCESSION AC022755  
VERSION AC022755.2 GI:9152928  
KEYWORDS HTG; HTGS PHASE0  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
1 (bases 1 to 92290)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
Homo sapiens, clone RP11-26007  
Unpublished  
2 (bases 1 to 92290)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,  
Boquelavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,  
Choeel,Y., Collangelo,M., Collins,S., Collymore,A., Cooke,P.,  
DeArrellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,  
Ferrela,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,  
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Holland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,  
McDonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,  
McPheters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,  
Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,  
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,  
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,  
Zimmer,A. and Zody,M.

TITLE  
JOURNAL  
COMMENT

Direct Submission  
Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jul 13, 2000 this sequence version replaced gi:6922579.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L5820  
Center clone name: 260\_O\_7  
-----

\* NOTE: This record contains 90 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 920: contig of 920 bp in length  
\* 921 1020: gap of 100 bp  
\* 1021 1939: contig of 919 bp in length  
\* 1940 2039: gap of 100 bp  
\* 2040 2949: contig of 910 bp in length  
\* 2950 3049: gap of 100 bp  
\* 3050 3977: contig of 928 bp in length  
\* 3978 4077: gap of 100 bp  
\* 4078 4982: contig of 905 bp in length  
\* 4983 5082: gap of 100 bp  
\* 5083 6030: contig of 948 bp in length  
\* 6031 6130: gap of 100 bp  
\* 6131 7112: contig of 982 bp in length  
\* 7113 7212: gap of 100 bp  
\* 7213 8147: contig of 935 bp in length  
\* 8148 8247: gap of 100 bp  
\* 8248 9168: contig of 921 bp in length  
\* 9169 9268: gap of 100 bp  
\* 9269 10218: contig of 950 bp in length  
\* 10219 10318: gap of 100 bp  
\* 10319 11273: contig of 955 bp in length  
\* 11274 11373: gap of 100 bp  
\* 11374 12290: contig of 917 bp in length  
\* 12291 12390: gap of 100 bp  
\* 12391 13336: contig of 946 bp in length  
\* 13337 13436: gap of 100 bp  
\* 13437 14393: contig of 957 bp in length  
\* 14394 14493: gap of 100 bp  
\* 14494 15426: contig of 933 bp in length  
\* 15427 15526: gap of 100 bp  
\* 15527 16456: contig of 930 bp in length  
\* 16457 16556: gap of 100 bp  
\* 16557 17499: contig of 943 bp in length  
\* 17500 17599: gap of 100 bp  
\* 17600 18532: contig of 933 bp in length  
\* 18533 18633: contig of 938 bp in length  
\* 18633 19670: gap of 100 bp  
\* 19671 20606: contig of 936 bp in length  
\* 20607 20706: gap of 100 bp  
\* 20707 21656: contig of 950 bp in length  
\* 21657 21756: gap of 100 bp  
\* 21757 22625: contig of 869 bp in length  
\* 22626 22725: gap of 100 bp  
\* 22726 23661: contig of 936 bp in length  
\* 23662 23761: gap of 100 bp

```
* 23762 24718: contig of 957 bp in length
* 24719 24818: gap of 100 bp
* 24819 25718: contig of 900 bp in length
* 25719 25818: gap of 100 bp
* 25819 26719: contig of 939 bp in length
* 26720 26758: gap of 100 bp
* 26759 27779: contig of 922 bp in length
* 27780 27879: gap of 100 bp
* 27880 28886: contig of 907 bp in length
* 28887 29843: contig of 957 bp in length
* 29844 29943: gap of 100 bp
* 29944 30893: contig of 949 bp in length
* 30893 31949: gap of 100 bp
* 31949 32049: contig of 957 bp in length
* 32050 32998: gap of 100 bp
* 32999 33614: contig of 516 bp in length
* 33615 33714: gap of 100 bp
* 33715 34650: contig of 936 bp in length
* 34651 34750: gap of 100 bp
* 34751 35657: contig of 907 bp in length
* 35658 35757: gap of 100 bp
* 35758 36709: contig of 952 bp in length
* 36710 36809: gap of 100 bp
* 36810 37749: contig of 940 bp in length
* 37750 37849: gap of 100 bp
* 37850 38783: contig of 934 bp in length
* 38784 38883: gap of 100 bp
* 38884 39794: contig of 911 bp in length
* 39795 39894: gap of 100 bp
* 39895 40848: contig of 954 bp in length
* 40849 41895: gap of 100 bp
* 41896 41995: contig of 947 bp in length
* 41996 42944: contig of 949 bp in length
* 42945 43044: gap of 100 bp
* 43045 43939: contig of 895 bp in length
* 43940 44039: gap of 100 bp
* 44040 44990: contig of 951 bp in length
* 44991 45090: gap of 100 bp
* 45091 46018: contig of 928 bp in length
* 46019 46118: gap of 100 bp
* 46119 47068: contig of 950 bp in length
* 47069 47169: gap of 100 bp
* 47170 48078: contig of 910 bp in length
* 48079 48178: gap of 100 bp
* 48179 49115: contig of 937 bp in length
* 49116 49215: gap of 100 bp
* 49216 50159: contig of 944 bp in length
* 50160 50259: gap of 100 bp
* 50260 51170: contig of 911 bp in length
* 51171 51270: gap of 100 bp
* 51271 52219: contig of 949 bp in length
* 52220 52319: gap of 100 bp
* 52320 53290: contig of 971 bp in length
* 53291 53390: gap of 100 bp
* 53391 54336: contig of 946 bp in length
* 54337 54436: gap of 100 bp
* 54437 55366: contig of 930 bp in length
* 55367 55466: gap of 100 bp
* 55467 56385: contig of 918 bp in length
* 56386 56484: gap of 100 bp
* 56485 57398: contig of 914 bp in length
* 57399 57499: gap of 100 bp
* 57500 58438: contig of 940 bp in length
* 58439 58539: gap of 100 bp
* 58540 59492: contig of 954 bp in length
* 59493 59592: gap of 100 bp
* 59593 60462: contig of 870 bp in length
* 60463 60563 61486: contig of 924 bp in length
```

```
* 61487 61586: gap of 100 bp
* 61587 62505: contig of 919 bp in length
* 62506 63548: contig of 943 bp in length
* 63549 64570: contig of 922 bp in length
* 64571 64670: gap of 100 bp
* 64671 65590: contig of 920 bp in length
* 65591 66626: contig of 936 bp in length
* 66627 66726: gap of 100 bp
* 66727 67621: contig of 895 bp in length
* 67622 67721: gap of 100 bp
* 67722 68680: contig of 958 bp in length
* 68681 69724: contig of 945 bp in length
* 69725 69824: gap of 100 bp
* 69825 70732: contig of 908 bp in length
* 70733 70832: gap of 100 bp
* 70833 71731: contig of 899 bp in length
* 71732 71831: gap of 100 bp
* 71832 72744: contig of 913 bp in length

Query Match 80.9%; Score 17.8; DB 14; Length 92290;
Best Local Similarity 90.5%; Pred. No. 3.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTAACCTGGA 22
DB 54900 CATGCTCACACATTAACCTGGA 54880

RESULT 31
AC114055_2
WPCOMMENT
Sequence split into 6 fragments LOCUS AC114055 Accession AC114055
Fragment Name Begin End
AC114055_0 1 110000
AC114055_1 100001 210000
AC114055_2 200001 310000
AC114055_3 300001 410000
AC114055_4 400001 510000
AC114055_5 500001 581269
Continuation (3 of 6) of AC114055 from base 200001 (AC114055 Rattus norvegicus clone CH.

Query Match 80.9%; Score 17.8; DB 14; Length 110000;
Best Local Similarity 90.5%; Pred. No. 3.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTAACCTGGA 22
DB 41630 CATGCTCACACATTAACCTGGA 41650

RESULT 32
AC128511_2/c
WPCOMMENT
Sequence split into 4 fragments LOCUS AC128511 Accession AC128511
Fragment Name Begin End
AC128511_0 1 110000
AC128511_1 100001 210000
AC128511_2 200001 310000
AC128511_3 300001 383149
Continuation (3 of 4) of AC128511 from base 200001 (AC128511 Rattus norvegicus clone CH.

Query Match 80.9%; Score 17.8; DB 14; Length 110000;
Best Local Similarity 90.5%; Pred. No. 3.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCATGCTCACACATTAACCTGG 21
DB 27222 TCATGCTCACACATTAACCTGG 27202
```

```

RESULT 33
AC119050/c
LOCUS      AC119050      130912 bp      DNA      linear      HTG 17-JUL-2002
DEFINITION  Gallus gallus clone WAG-39H1, WORKING DRAFT SEQUENCE.
ACCESSION  AC119050
VERSION    AC119050.1  GI:20279416
KEYWORDS   HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE     Gallus gallus (chicken)
ORGANISM   Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE  1 (bases 1 to 130912)
AUTHORS   Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,
            Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
            Haghighi,P., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E., Laric,P.,
            Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
            Margulies,E.H., Masello,C., Maskeri,B., Mastrian,S.D.,
            McCloskey,J.C., McDowell,J., Peguirigan,C., Pearson,R.,
            Portnoy,M.E., Prasad,A., Schueler,M.G., Stantropop,S., Thomas,J.W.,
            Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
            Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
            NISC Comparative Sequencing Initiative
            Unpublished
            2 (bases 1 to 130912)
            Green,E.D.
            Direct Submission
            Submitted (24-APR-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
            3 (bases 1 to 130912)
            Green,E.D.
            Direct Submission
            Submitted (17-JUL-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
            ----- Genome Center
            Center: NIH Intramural Sequencing Center
            Center code: NISC
            Web site: http://www.nisc.nih.gov
            Contact: nisc.zoo@nhgri.nih.gov
            ----- Project Information
            Center project name: apr
            Center clone name: 039401
            ----- Summary Statistics
            Sequencing vector: plasmid; n/a; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 129708 bases at least Q40
            Consensus quality: 130306 bases at least Q30
            Consensus quality: 130641 bases at least Q20
            Insert size: 114000; agarose-fp
            Insert size: 130912; sum-of-contigs
            Quality coverage: 7.30x in Q20 bases; agarose-fp
            Quality coverage: 6.36x in Q20 bases; sum-of-contigs
            -----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            * 1 130912: contig of 130912 bp in length.
            FEATURES
            source      Location/Qualifiers
            1..130912
            /organism="Gallus gallus"
            /mol_type="genomic DNA"
            /db_xref="taxon:9031"
            /clone="WAG-39H1"
            /clone_lib="WAG"
            misc_feature 1..130912

AC119050      130912 bp      DNA      linear      HTG 17-JUL-2002
Gallus gallus clone WAG-39H1, WORKING DRAFT SEQUENCE.
AC119050
AC119050.1  GI:20279416
HTG; HTGS PHASE1; HTGS DRAFT.
Gallus gallus (chicken)
Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 130912)
Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,
Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
Haghighi,P., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E., Laric,P.,
Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
Margulies,E.H., Masello,C., Maskeri,B., Mastrian,S.D.,
McCloskey,J.C., McDowell,J., Peguirigan,C., Pearson,R.,
Portnoy,M.E., Prasad,A., Schueler,M.G., Stantropop,S., Thomas,J.W.,
Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 130912)
Green,E.D.
Direct Submission
Submitted (24-APR-2002) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 130912)
Green,E.D.
Direct Submission
Submitted (17-JUL-2002) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.zoo@nhgri.nih.gov
----- Project Information
Center project name: apr
Center clone name: 039401
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 129708 bases at least Q40
Consensus quality: 130306 bases at least Q30
Consensus quality: 130641 bases at least Q20
Insert size: 114000; agarose-fp
Insert size: 130912; sum-of-contigs
Quality coverage: 7.30x in Q20 bases; agarose-fp
Quality coverage: 6.36x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 130912: contig of 130912 bp in length.
FEATURES
source      Location/Qualifiers
1..130912
/organism="Gallus gallus"
/mol_type="genomic DNA"
/db_xref="taxon:9031"
/clone="WAG-39H1"
/clone_lib="WAG"
misc_feature 1..130912

AC119050      130912 bp      DNA      linear      PRI 18-MAY-2005
Human DNA sequence from clone RP11-215C13 on chromosome 10,
complete sequence.
AL731573      131465 bp      DNA      linear      PRI 18-MAY-2005
AL731573      AC012559
AL731573.8  GI:20428967
HTG.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homnidae; Homo.
            1 (bases 1 to 131465)
            Grafham,D.
            Direct Submission
            Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
            On May 3, 2002 this sequence version replaced gi:20386962.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep
            This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 10, constructed by the Sanger Centre Chromosome 10
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr10
            ----- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vegas@sanger.ac.uk
            -----
            RP11-215C13 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            Draft Sequence Produced by Genome Therapeutics Corp, 100 Beaver
            Street, Waltham, MA 02453, USA
            http://www.genomecorp.com
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
            FEATURES
            Location/Qualifiers
            1..131465
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="10"
            /clone="RP11-215C13"
            /clone_lib="RPCI-11.1"
            2000
            /note="Clone_right_end: RP11-538014"
            misc_feature 2000

```



```

misc_feature 129466
/note="clone_left_end: RP11-308N23"

ORIGIN
Query Match      80.9%  Score 17.8;  DB 8;  Length 131465;
Best Local Similarity 90.5%  Pred.No. 3.1e+02;
Matches 19;  Conservative 0;  Mismatches 2;  Indels 0;  Gaps 0;

Qy  2  CATGCTCACATTAACCTGGA 22
      |||||  |||||  |||||
Db  65418  CATGCTCAATCAATTAACCTGGA 65438

RESULT 35
AC136548/c
LOCUS
DEFINITION
AC136548      140414 bp      DNA      linear      HTG 06-NOV-2002
Rattus norvegicus clone CH230-154H13, *** SEQUENCING IN PROGRESS
***, 66 unordered pieces.
AC136548
VERSION
AC136548.1  GI:24635346
KEYWORDS
HTG; HTGS PHASE1.
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 140414)
Muzny,D,Marie., Metzker,M, Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsebrooks,S., Amin,A., Angiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Blawid,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.W., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Georgiev,G., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M.,
Hollins,B., Howells,S., Hulvik,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,S., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Loulsegad,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Napua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M., McNeill,T., Meenen,E., Milosavljevic,A.,
Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K.,
Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D.,
Newton,N., Nguyen,N., Norris,S., Nwaokeleneh,O., Okwuonu,G.,
Olarunpugoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H.,
Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A.,
Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E.,
Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y.,
Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A.,
Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S.,
Scott,G., Shatman,S., Shen,H., Shetty,J., Shvartsbeyn,A.,
Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E.,
Song,X.-Z., Sorrelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A.,
Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S.,
Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villanueva,D.,
Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J.,
Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczky,R.,
Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S.,
Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X.,

```

```

Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R.,
Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 140414)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (06-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: KDHG
Center clone name: CH230-154H13
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 83751 bases at least Q40
Consensus quality: 88044 bases at least Q30
Consensus quality: 90796 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 66 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 1210: contig of 1210 bp in length
* 1211 1310: gap of unknown length
* 1311 2505: contig of 1195 bp in length
* 2506 2605: gap of unknown length
* 2606 4065: contig of 1460 bp in length
* 4066 4165: gap of unknown length
* 4166 5655: contig of 1489 bp in length
* 5655 5754: gap of unknown length
* 5755 6897: contig of 1143 bp in length
* 6898 6997: gap of unknown length
* 6998 8045: contig of 1048 bp in length
* 8046 8146: gap of unknown length
* 8146 9167: contig of 1021 bp in length
* 9167 9267: gap of unknown length
* 9267 10740: contig of 1474 bp in length
* 10741 10840: gap of unknown length
* 10841 12163: contig of 1323 bp in length
* 12164 12263: gap of unknown length
* 12264 13445: contig of 1182 bp in length
* 13445 13545: gap of unknown length
* 13546 14737: contig of 1192 bp in length
* 14738 14837: gap of unknown length
* 14838 16002: contig of 1165 bp in length
* 16003 16102: gap of unknown length
* 16103 17271: contig of 1169 bp in length
* 17271 17371: gap of unknown length
* 17372 18498: contig of 1127 bp in length
* 18499 18598: gap of unknown length
* 18599 19627: contig of 1029 bp in length
* 19628 19727: gap of unknown length
* 19728 21347: contig of 1520 bp in length
* 21348 23071: gap of unknown length
* 23072 23171: contig of 1724 bp in length
* 23172 24712: contig of 1541 bp in length
* 24713 26537: contig of unknown length
* 24813

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

* 26538	26637:	gap of unknown length			
* 26638	28039:	contig of 1402 bp in length			
* 28040	28139:	gap of unknown length			
* 28140	29149:	contig of 1010 bp in length			
* 29150	29249:	gap of unknown length			
* 29250	30393:	contig of 1144 bp in length			
* 30394	30493:	gap of unknown length			
* 30494	32159:	contig of 1666 bp in length			
* 32160	32259:	gap of unknown length			
* 32260	33897:	contig of 1638 bp in length			
* 33898	33997:	gap of unknown length			
* 33998	35759:	contig of 1762 bp in length			
* 35760	35859:	gap of unknown length			
* 35860	37200:	contig of 1341 bp in length			
* 37201	37300:	gap of unknown length			
* 37301	39183:	contig of 1883 bp in length			
* 39184	39283:	gap of unknown length			
* 39284	41348:	contig of 2065 bp in length			
* 41349	41448:	gap of unknown length			
* 41449	42728:	contig of 1280 bp in length			
* 42729	42828:	gap of unknown length			
* 42829	44460:	contig of 1632 bp in length			
* 44461	44560:	gap of unknown length			
* 44561	46807:	contig of 2247 bp in length			
* 46808	46907:	gap of unknown length			
* 46908	48733:	contig of 1846 bp in length			
* 48754	48833:	gap of unknown length			
* 48854	50606:	contig of 1753 bp in length			
* 50607	50706:	gap of unknown length			
* 50707	52427:	contig of 1721 bp in length			
* 52428	52527:	gap of unknown length			
* 52528	54354:	contig of 1827 bp in length			
* 54355	54454:	gap of unknown length			
* 54455	56599:	contig of 2145 bp in length			
* 56600	56699:	gap of unknown length			
* 56700	58435:	contig of 1736 bp in length			
* 58436	58535:	gap of unknown length			
* 58536	60799:	contig of 2264 bp in length			
* 60800	60899:	gap of unknown length			
* 60900	62933:	contig of 2034 bp in length			
* 62934	63033:	gap of unknown length			
* 63034	64121:	contig of 1088 bp in length			
* 64122	64221:	gap of unknown length			
* 64222	66561:	contig of 2340 bp in length			
* 66562	66661:	gap of unknown length			
* 66662	68500:	contig of 1839 bp in length			
* 68501	68600:	gap of unknown length			
* 68601	70239:	contig of 1639 bp in length			
* 70240	70339:	gap of unknown length			
* 70340	72719:	contig of 2380 bp in length			
* 72720	72819:	gap of unknown length			
* 72820	74704:	contig of 1885 bp in length			
* 74705	74804:	gap of unknown length			
* 74805	77593:	contig of 2789 bp in length			
* 77594	77693:	gap of unknown length			
* 77694	79780:	contig of 2087 bp in length			
* 79781	79880:	gap of unknown length			
* 79881	82264:	contig of 2384 bp in length			
* 82265	82364:	gap of unknown length			
* 82365	84436:	contig of 2072 bp in length			
* 84437	84536:	gap of unknown length			
* 84537	86599:	contig of 2059 bp in length			
* 86599	86695:	gap of unknown length			
* 86696	88875:	contig of 2180 bp in length			
* 88876	88975:	gap of unknown length			
* 88976	91275:	contig of 2300 bp in length			
* 91276	93175:	gap of unknown length			
* 93176	93913:	contig of 2538 bp in length			
* 93914	94013:	gap of unknown length			

Query Match 80.9%; Score 17.8; DB 14; Length 140414;  
Best Local Similarity 90.5%; Pred. No. 3.1e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY

Db

RESULT 36

ALI39812/c

LOCUS

DEFINITION

Human DNA sequence from clone RPI-19N1 on chromosome Xq21.33-22.3

Contains the gene for a novel protein similar to chondroitin

sulfate GALNACT-2 (GALNACT-2), complete sequence.

ACCESSION

ALI39812

VERSION

HTG; GALNACT-2

KEYWORDS

Source

ORGANISM

Homo sapiens (human)

REFERENCE

1 (bases 1 to 144242)

Grafham,D.

Direct Submission

Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk

Clone requests: clonerequest@sanger.ac.uk

On Jul 14, 2000 this sequence version replaced gi:8918976.

The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:

Em: EMBL; Swi: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information

on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence

was generated from part of bacterial clone contigs of human

chromosome X, constructed by the Sanger Centre Chromosome X Mapping

Group. Further information can be found at

http://www.sanger.ac.uk/HGP/ChrX

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: vegas@sanger.ac.uk

-----

This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate

chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by at least

one subclone; and the assembly was confirmed by restriction digest,

except on the rare occasion of the clone being a YAC.

RPI-19N1 is from the library RPI-1 constructed by the group of

Pieter de Jong. For further details see

http://www.chori.org/bacpac/home.htm

VECTOR: pCYPAC2.

Location/Qualifiers

1. .144242

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/map="q21.33-22.3"

/clone="RPI-19N1"

/clone\_lib="RPI-1"

-33224\_..35282

/locus\_tag="RPI-19N1.1-001"

-33224\_..35282

/product="novel protein similar to chondroitin sulfate

GALNACT-2 (GALNACT-2)"

/note="match: ESTs: AA043195.1 AA043597.1 AA079045.1

AA281222.1 AA410548.1 AA715822.1 AA721408.1 AA736590.1

AA748617.1 AA773203.1 AA825719.1 AF063526.1 AF110652.1

AF127063.1 AF282011.1 AF356844.1 AF640527.1 AF658628.1

AF673813.1 AF680304.1 AF689553.1 AF699487.1 AF913223.1

AF922789.1 AF933701.1 AF990225.1 AF049045 AF049046

1 TCATGCTCACACATTAACTGG 21

||||| ||||| ||||| |||||

72 TCATGTCACACATTCATCG 52

ALI39812 144242 bp DNA linear PRI 18-MAY-2005

Human DNA sequence from clone RPI-19N1 on chromosome Xq21.33-22.3

Contains the gene for a novel protein similar to chondroitin

sulfate GALNACT-2 (GALNACT-2), complete sequence.

ALI39812

ALI39812.11 GI:9187185

HTG; GALNACT-2

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

1 (bases 1 to 144242)

Grafham,D.

Direct Submission

Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk

Clone requests: clonerequest@sanger.ac.uk

On Jul 14, 2000 this sequence version replaced gi:8918976.

The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:

Em: EMBL; Swi: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information

on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence

was generated from part of bacterial clone contigs of human

chromosome X, constructed by the Sanger Centre Chromosome X Mapping

Group. Further information can be found at

http://www.sanger.ac.uk/HGP/ChrX

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: vegas@sanger.ac.uk

-----

This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate

chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by at least

one subclone; and the assembly was confirmed by restriction digest,

except on the rare occasion of the clone being a YAC.

RPI-19N1 is from the library RPI-1 constructed by the group of

Pieter de Jong. For further details see

http://www.chori.org/bacpac/home.htm

VECTOR: pCYPAC2.

Location/Qualifiers

1. .144242

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/map="q21.33-22.3"

/clone="RPI-19N1"

/clone\_lib="RPI-1"

-33224\_..35282

/locus\_tag="RPI-19N1.1-001"

-33224\_..35282

/product="novel protein similar to chondroitin sulfate

GALNACT-2 (GALNACT-2)"

/note="match: ESTs: AA043195.1 AA043597.1 AA079045.1

AA281222.1 AA410548.1 AA715822.1 AA721408.1 AA736590.1

AA748617.1 AA773203.1 AA825719.1 AF063526.1 AF110652.1

AF127063.1 AF282011.1 AF356844.1 AF640527.1 AF658628.1

AF673813.1 AF680304.1 AF689553.1 AF699487.1 AF913223.1

AF922789.1 AF933701.1 AF990225.1 AF049045 AF049046

AV341861 AV343539 AV372895 AW007673.1 AW025693.1  
 AW026446.1 AW088009.1 AW264439.1 AW440954.1 AW471241.1  
 AW581342.1 C01580 D61822 D62659 D62752 D79651 D82671  
 H06061.1 H13398.1 245941  
 match: cDNAs: AB079252.1 AB090811.1 AF116646.1 AK002126.1  
 AK074474.1 AL157483.1 BC030288.1"  
 <33224..33516  
 /locus\_tag="RP1-19N1.1-001"  
 /standard\_name="OTTHUMP0000023782"  
 /notes="match: proteins: Q8C1F4 Q8N6G5 Q8R5A2 Q8TCH4 Q8MWT2  
 Q9D2M1 Q9H518 Q9P116"  
 /codon\_start=3  
 /product="novel protein similar to chondroitin sulfate  
 GALNACT-2 (GALNACT-2)"  
 /protein\_id="CAI42420.1"  
 /db\_xref="GI:57208784"  
 /db\_xref="UniProt/TREMBL:O5JUV9"  
 /translation="GFDMEVKGWGEDVHLYRKLRGDLIVRNVPGLFHLWHEKHC  
 ADLTPQRMCIQSKAMDEASKHGLMIVFREIEIETHLHKQAYRTNSEAVG"  
 33670..33675  
 /locus\_tag="RP1-19N1.1-001"  
 33745  
 /locus\_tag="RP1-19N1.1-001"  
 35255..35260  
 /locus\_tag="RP1-19N1.1-001"  
 35282  
 /locus\_tag="RP1-19N1.1-001"

polyA\_signal  
 polyA\_site  
 polyA\_signal  
 polyA\_site

ORIGIN  
 Query Match 80.9%; Score 17.8; DB 8; Length 144242;  
 Best Local Similarity 90.5%; Pred. No. 3.1e+02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTAACTGGA 22  
 |||||  
 Db 94298 CATGCTCACACATTCTTGA 94278  
 |||||

RESULT 37  
 CR847900/c  
 LOCUS 145618 bp DNA linear HTG 14-JUL-2005  
 DEFINITION Danio rerio chromosome 6 clone CH211-180C7, WORKING DRAFT SEQUENCE,  
 3 unordered pieces.  
 ACCESSION  
 VERSION CR847900.2 GI:70907993  
 KEYWORDS HTG: HTGS PHASE1: HTGS DRAFT; HTGS\_FULLTOP.  
 SOURCE Danio rerio (zebrafish)  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
 Cypriniformes; Cyprinidae; Danio.  
 1 (bases 1 to 145618)  
 Mclaren,S.  
 Direct Submission  
 Submitted (13-JUL-2005) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 zfish-help@sanger.ac.uk Clone requests:  
 http://www.sanger.ac.uk/Projects/D\_rerio/fags.shtml#dataeight  
 On Jul 15, 2005 this sequence version replaced gi:53850236.  
 ----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: zfish-help@sanger.ac.uk  
 ----- Project Information  
 Center project name: zc180C7  
 ----- Summary Statistics  
 Assembly program: XGAP4; version 4.5  
 Chemistry: Dye-terminator; 100% of reads  
 Consensus quality: 144858 bases at least Q40  
 Consensus quality: 145001 bases at least Q30  
 Consensus quality: 145086 bases at least Q20  
 Insert size: 145418; sum-of-Contigs

Insert size: 146454; 5.7% error; agarose-fp  
 Quality coverage: 7.53x in Q20 bases; sum-of-Contigs Quality  
 coverage: 7.48x in Q20 bases; agarose-fp  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 3 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 25402: contig of 25402 bp in length  
 \* 25403 25502: gap of 100 bp  
 \* 25503 75035: contig of 49533 bp in length  
 \* 75036 75135: gap of 100 bp  
 \* 75136 145618: contig of 70483 bp in length.

FEATURES  
 Location/Qualifiers  
 1..145618  
 /organism="Danio rerio"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7955"  
 /chromosome="6"  
 /clone="CH211-180C7"  
 /clone\_lib="CHORI-211"  
 1..25402  
 /note="assembly\_fragment:00007  
 fragment\_chain:1  
 clone\_end:17  
 vector\_side:left"  
 25503..75035  
 /note="assembly\_fragment:00282  
 fragment\_chain:1  
 clone\_end:SP6  
 vector\_side:right"  
 75136..145618  
 /note="assembly\_fragment:00891  
 fragment\_chain:1"

misc\_feature  
 misc\_feature  
 misc\_feature

ORIGIN  
 Query Match 80.9%; Score 17.8; DB 14; Length 145618;  
 Best Local Similarity 90.5%; Pred. No. 3.1e+02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCATGCTCACACATTAACTGG 21  
 |||||  
 Db 14542 TCATGTTACACATTAACTGG 14522  
 |||||

RESULT 38  
 BX649515/c  
 LOCUS 153400 bp DNA linear VRT 23-APR-2004  
 DEFINITION Zebrafish DNA sequence from clone CH211-120F15 in linkage group 3,  
 complete sequence.  
 ACCESSION BX649515  
 VERSION BX649515.7 GI:46559237  
 KEYWORDS HTG.  
 SOURCE Danio rerio (zebrafish)  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
 Cypriniformes; Cyprinidae; Danio.  
 1 (bases 1 to 153400)  
 Pelan,S.  
 Direct Submission  
 Submitted (23-APR-2004) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
 On Apr 23, 2004 this sequence version replaced gi:46016481.  
 ----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: http://www.sanger.ac.uk

Contact: zfish-help@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) Clone-derived Zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.

Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhirong Bao and Sean Eddy, submitted), and those beginning 'drv' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see

[http://www.sanger.ac.uk/Projects/D\\_reio/fishmask.shtml](http://www.sanger.ac.uk/Projects/D_reio/fishmask.shtml)

CH211-120F15 is from a CHORI-211 BAC library

VECTOR: pTARBAC2.1.

```

FEATURES
    source
        1..153400
            /organism="Danio rerio"
            /mol_type="genomic DNA"
            /db_xref="taxon:7955"
            /clone="CH211-120F15"
            /clone_lib="CHORI-211"

```

#### ORIGIN

```

Query Match      80.9%; Score 17.8; DB 5; Length 153400;
Best Local Similarity 90.5%; Pred. No. 3.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      2  CATGCTCACACATTAACTGGA 22
        |||||
Db      58498 CATGCTCACACATTAGCTTGA 58478

```

```

RESULT 39
BX072552/c
LOCUS    BX072552      160230 bp      DNA      linear      ROD 27-JUN-2003
DEFINITION
    Mouse DNA sequence from clone RP23-147124 on chromosome X, complete
    sequence.
ACCESSION
    BX072552
VERSION   BX072552.10 GI:32328860
KEYWORDS  HTG.
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus

```

```

REFERENCE
AUTHORS   Chapman,J.
TITLE      Direct Submission
JOURNAL    Submitted (26-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,
           Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
           humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT    On Jun 27, 2003 this sequence version replaced gi:32187941.
           Sequence from the Mouse Genome Sequencing Consortium whole genome
           shotgun may have been used to confirm this sequence. Sequence data
           from the whole genome shotgun alone has only been used where it has

```

a phred quality of at least 30.

----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: <http://www.sanger.ac.uk>  
 Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at

[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) RP23-147124 is from the RPCI-23 Mouse BAC Library constructed by the group of Pieter de Jong.

For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACe3.6.

```

FEATURES
    Location/Qualifiers
        1..160230
            /organism="Mus musculus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10090"
            /chromosome="X"
            /clone="RP23-147124"
            /clone_lib="RPCI-23"

```

#### ORIGIN

```

Query Match      80.9%; Score 17.8; DB 9; Length 160230;
Best Local Similarity 90.5%; Pred. No. 3.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      2  CATGCTCACACATTAACTGGA 22
        |||||
Db      136184 CATGCTCACACACCAACTGGA 136164

```

#### RESULT 40

```

AC116927
LOCUS    AC116927      166937 bp      DNA      linear      HTG 19-OCT-2002
DEFINITION
    Rattus norvegicus strain Brown Norway clone RP31-496H14, WORKING
    DRAFT SEQUENCE, 4 ordered pieces.
AC116927

```

```

ACCESSION
    AC116927.3 GI:24137421
VERSION   HTG; HTGS_PHASE2; HTGS_DRAFT.
KEYWORDS  Rattus norvegicus (Norway rat)
SOURCE    Rattus norvegicus
ORGANISM  Rattus norvegicus

```

```

REFERENCE
AUTHORS   Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
           Benjamen,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
           Carriaga,K., Coleman,B., Dietrich,N.L., Engle,J., Granite,S.,
           Guan,X., Gupta,J., Haghighi,P., Han,J., Hansen,N., Ho,S.-L.,
           Idol,J.R., Karlins,E., Laric,P., Lee-Lin,S.-O., Legaspi,R.,
           Maduro,O.L., Maduro,V.B., Margulies,E.H., Masiello,C., Maskeri,B.,
           Mastrian,S.D., McCloskey,J.C., McDowell,J., Pagnirigan,C.,
           Pearson,R., Portnoy,M.E., Prasad,A., Reddix-Dugue,N., Schandler,K.,
           Schueler,M.G., Sison,C., Stantripop,S., Thomas,J.W., Thomas,P.J.,
           Touchman,J.W., Vogt,J.L., Walker,M., Wetherby,K.D., Wiggins,L.,

```

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Young, A., Zhang, L.-H. and Green, E.D.  
NISC Comparative Sequencing Initiative  
Unpublished  
2 (bases 1 to 166937)  
Green, E.D.  
Direct Submission  
Submitted (03-APR-2002) NIH Intramural Sequencing Center, 8717  
Government Circle, Gaithersburg, MD 20877, USA  
3 (bases 1 to 166937)  
Green, E.D.  
Direct Submission  
Submitted (19-OCT-2002) NIH Intramural Sequencing Center, 8717  
Government Circle, Gaithersburg, MD 20877, USA  
On Oct 19, 2002 this sequence version replaced gi:23477839.  
----- Genome Center  
Center: NIH Intramural Sequencing Center  
Center code: NISC  
Web site: <http://www.nisc.nih.gov>  
Contact: [nisc\\_zoo@ngri.nih.gov](mailto:nisc_zoo@ngri.nih.gov)  
----- Project Information  
Center project name: cxs  
Center clone name: 496H14

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics  
Sequencing vector: plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 166108 bases at least Q40  
Consensus quality: 166312 bases at least Q30  
Consensus quality: 166493 bases at least Q20  
Insert size: 192000; agarose-fp  
Insert size: 166637; sum-of-contigs  
Quality coverage: 10.35x in Q20 bases; agarose-fp  
Quality coverage: 11.93x in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submittor.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

\* 1 4883: contig of 4883 bp in length  
\* 4884 4983: gap of unknown length  
\* 4984 80899: contig of 75916 bp in length  
\* 80900 80999: gap of unknown length  
\* 81000 133877: contig of 52878 bp in length  
\* 133878 133977: gap of unknown length  
\* 133978 166937: contig of 32960 bp in length.

FEATURES  
source  
1. .166937  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/strain="Brown Norway"  
/db\_xref="taxon:10116"  
/clone="RP31-496H14"  
/clone\_lib="RP31"  
1. .4883  
/note="assembly\_fragment"

clone\_end:SP6  
vector\_side:left"  
4884. .4983  
/estimated\_length=unknown  
4884. .80899  
/note="assembly\_fragment"  
80900. .80999  
/estimated\_length=unknown  
81000. .133877  
/note="assembly\_fragment"  
133878. .133977  
/estimated\_length=unknown  
133978. .166937  
/note="assembly\_fragment"  
clone\_end:T7  
vector\_side:right"

gap  
misc\_feature  
gap  
misc\_feature  
gap  
misc\_feature

## ORIGIN

Query Match 80.9%; Score 17.8; DB 14; Length 166937;  
Best Local Similarity 90.5%; Pred. NO. 3.1e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TCATGCTCACACATTTAACTGG 21  
|||  
Db 103484 TCCTGCTCACACTAACTGG 103504

Search completed: November 20, 2005, 18:34:11  
Job time : 811.309 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2005, 15:03:55 ; Search time 357.432 Seconds  
(without alignments)  
410.213 Million cell updates/sec

Title: US-10-627-757-28

Perfect score: 22  
Sequence: 1 tcatgctcacattactgga 22

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- N\_Geneseq\_21.\*  
1: Geneseqn1980s.\*  
2: Geneseqn1990s.\*  
3: Geneseqn2000s.\*  
4: Geneseqn2001as.\*  
5: Geneseqn2001bs.\*  
6: Geneseqn2002as.\*  
7: Geneseqn2002bs.\*  
8: Geneseqn2003as.\*  
9: Geneseqn2003bs.\*  
10: Geneseqn2003cs.\*  
11: Geneseqn2003ds.\*  
12: Geneseqn2004as.\*  
13: Geneseqn2004bs.\*  
14: Geneseqn2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	22	100.0	22	ADL14976	ADL14976 Human gla
2	22	100.0	22	ADW86598	ADW86598 PCR prime
3	22	100.0	1116	ADL14956	ADL14956 Human gla
4	22	100.0	1116	ADW86578	ADW86578 Human gla
5	22	100.0	46951	ADL13891	ADL13891 Human opt
6	19.4	88.2	731	ACH89765	ACH89765 Human gen
7	19.4	88.2	4557	AAS92423	AAS92423 DNA encod
8	19.4	88.2	8277	ADK15078	ADK15078 Urinary s
9	17.8	80.9	842	ADK61760	ADK61760 Plant ful
10	17.2	78.2	110000	ABD32968_2	Continuation (3 of
11	16.8	76.4	27251	ADZ59508	ADZ59508 Secondary
12	16.8	76.4	110000	ACN44934_2	Continuation (3 of
13	16.8	76.4	110000	ADZ46976_4	Continuation (5 of
14	16.4	74.5	288	ABN18239	ABN18239 Human ORF
15	16.4	74.5	15845	ADM68759	ADM68759 Myzus per
16	16.4	74.5	247544	ADQ59419	ADQ59419 Human can
17	16.4	74.5	247654	ADZ13712	ADZ13712 Murine ca
18	16.2	73.6	65	ADG89121	ADG89121 Cancer de
19	16.2	73.6	210	ADL12599	ADL12599 Mutant hu

20	16.2	73.6	390	12	ADL12627	ADL12627 Human BRC
21	16.2	73.6	456	9	ACH48346	ACH48346 Human lun
22	16.2	73.6	1083	4	AAD09872	AAD09872 Infectiou
23	16.2	73.6	1166	4	AAD09875	AAD09875 Infectiou
24	16.2	73.6	1167	4	AAD09874	AAD09874 Infectiou
25	16.2	73.6	1272	8	ACA21535	ACA21535 Prokaryot
26	16.2	73.6	1320	4	ACA288168	ACA288168 Infectiou
27	16.2	73.6	1320	4	AAC88169	AAC88169 Infectiou
28	16.2	73.6	2325	13	ADK45527	ADK45527 Streptoco
29	16.2	73.6	2343	13	ADR92735	ADR92735 Novel S.
30	16.2	73.6	2343	14	AEA56605	Aea56605 Streptoco
31	16.2	73.6	2352	6	ABN68555	ABN68555 Streptoco
32	16.2	73.6	2374	3	AAZ97024	AAZ97024 Human sec
33	16.2	73.6	2374	9	ACH66653	ACH66653 Novel hum
34	16.2	73.6	2576	4	ABL25820	ABL25820 Drosophil
35	16.2	73.6	3404	14	AEA61096	Aea61096 Human GUC
36	16.2	73.6	3559	12	ADJ96274	ADJ96274 Human bre
37	16.2	73.6	3559	14	ADX06107	ADX06107 Cyclinde
38	16.2	73.6	3559	14	ADW98062	ADW98062 Human bre
39	16.2	73.6	3682	12	ADK67836	ADK67836 Human BRC
40	16.2	73.6	3682	12	ADJ96279	ADJ96279 Human bre
41	16.2	73.6	3682	14	ADW98067	ADW98067 Human bre
42	16.2	73.6	3796	12	ADJ96280	ADJ96280 Human bre
43	16.2	73.6	3796	14	ADW98068	ADW98068 Human bre
44	16.2	73.6	4401	3	AAZ98404	AAZ98404 Rhesus mo
45	16.2	73.6	4401	6	ABK40736	ABK40736 Monkey be

ALIGNMENTS

RESULT 1

ADL14976  
ID ADL14976 standard; DNA; 22 BP.

AC ADL14976;

XX 06-MAY-2004 (first entry)

XX Human glaucoma-related optineurin (OPTN) exon 10 PCR primer SR10.

XX Human; glaucoma; optineurin; OPTN; diagnosis; PCR; primer; ss.

XX Homo sapiens.

XX EP1388590-A2.

XX 11-FEB-2004.

XX 29-JUL-2003; 2003EP-00447201.

XX 02-AUG-2002; 2002JP-00226612.

XX (SYSM-) SYSMEX CORP.

XX Kouchi Y, Masago A, Takahata T;

XX WPI; 2004-146134/15.

PT Gene assay for predicting future onset of glaucoma, particularly primary  
open angle glaucoma or normal ocular tension glaucoma, comprises  
detecting a mutation of at least one base of the optineurin gene.

PS Claim 9; SEQ ID NO 28; 31pp; English.

XX The present sequence is that of PCR primer SR10 for exon 10 ADL14956 of  
the glaucoma-associated gene, OPTN (optineurin) ADL14949. The invention  
relates to a gene assay method for predicting future onset of primary  
open angle glaucoma and/or normal ocular tension glaucoma. This involves  
detecting a mutation in the OPTN gene coding sequence, specifically a  
substitution of G for A at position 619 and/or a substitution of A for G  
at position 898 of the OPTN coding sequence. The mutation(s) is detected  
using a nucleic acid amplification method using primers specific for the

```

CC different exons of the coding sequence, including primers SR10 and SF10
CC ADL14975 for exon 10.
XX
SQ Sequence 22 BP; 7 A; 6 C; 3 G; 6 T; 0 U; 0 Other;

Query Match      100.0%; Score 22; DB 12; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.45;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCATGCTCACATTAACTGGA 22
   |||||
Db 1 TCATGCTCACATTAACTGGA 22

RESULT 2
ADW86598
ID ADW86598 standard; DNA; 22 BP.
XX
AC ADW86598;
XX
XX 21-APR-2005 (first entry)
XX
XX PCR primer used to amplify human optineurin (OPTN) gene exon 10 Seq 28.
XX
XX glaucoma; optineurin; ophthalmological; ss; ocular disease;
XX DNA amplification; genetic marker; PCR; primer.
XX
XX Homo sapiens.
XX
XX JP2005034112-A.
XX
XX 10-FEB-2005.
XX
XX 29-JUL-2003; 2003JP-00281897.
XX
XX 02-AUG-2002; 2002JP-00226612.
XX
XX 30-JUN-2003; 2003JP-00188070.
XX
XX (TOAI-) TOA IYO DENSHI KK.
XX
XX WPI; 2005-156038/17.
XX
XX Estimating risk of onset of glaucoma, involves analyzing mutation in any
XX one portion of Optic new phosphorous gene, and utilizing analyzed
XX mutation as index for estimating risk of onset of glaucoma.
XX
XX Example 1; SEQ ID NO 28; 13pp; Japanese.
XX
XX This invention relates to a novel method for estimating the risk of onset
XX of glaucoma. Specifically, it refers to a method that involves analyzing
XX a mutation in any one portion of an optineurin (OPTN) gene and utilizing
XX the analyzed mutation as an index for estimating the risk of onset of
XX glaucoma. The present invention describes oligonucleotides to detect
XX mutations that hybridize with one or more portions of the OPTN glaucoma
XX related gene. Accordingly, it provides oligos that detect A619G and G898A
XX mutations in the human OPTN gene of patients at risk of developing
XX glaucoma, in particular primary open-angle glaucoma and/or normal tension
XX glaucoma. These mutations will not be present in a sample obtained from a
XX healthy person who is not at risk of onset of glaucoma. This
XX oligonucleotide sequence is a PCR primer used to amplify a human OPTN
XX exon of the invention.
XX
SQ Sequence 22 BP; 7 A; 6 C; 3 G; 6 T; 0 U; 0 Other;

Query Match      100.0%; Score 22; DB 14; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.45;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCATGCTCACATTAACTGGA 22
   |||||
Db 1 TCATGCTCACATTAACTGGA 22

us-10-627-757-28.rng
RESULT 3
ADL14956/c
ID ADL14956 standard; DNA; 1116 BP.
XX
AC ADL14956;
XX
XX 06-MAY-2004 (first entry)
XX
XX Human glaucoma-related optineurin (OPTN) exon 10.
XX
XX Human; glaucoma; optineurin; OPTN; diagnosis; ds.
XX
XX Homo sapiens.
XX
XX Key location/Qualifiers
XX primer_bind complement(251..269)
XX /*tag= a
XX /*note= "Primer SF10"
XX FT exon 501..616
XX /*tag= b
XX FT /number= 10
XX primer_bind 765..786
XX /*tag= c
XX /*note= "Primer SR10"
XX
XX EP1388590-A2.
XX
XX 11-FEB-2004.
XX
XX 29-JUL-2003; 2003EP-00447201.
XX
XX 02-AUG-2002; 2002JP-00226612.
XX
XX (SYSM-) SYSMEX CORP.
XX
XX Kouchi Y, Masago A, Takahata T;
XX WPI; 2004-146134/15.
XX
XX Gene assay for predicting future onset of glaucoma, particularly primary
XX open angle glaucoma or normal ocular tension glaucoma, comprises
XX detecting a mutation of at least one base of the optineurin gene.
XX
XX Example 1; SEQ ID NO 8; 31pp; English.
XX
XX The present sequence comprises exon 10 of the glaucoma-associated gene,
XX OPTN (optineurin) ADL14949. The invention relates to a gene assay method
XX for predicting future onset of primary open angle glaucoma and/or normal
XX ocular tension glaucoma. This involves detecting a mutation in the OPTN
XX gene coding sequence, specifically a substitution of G for A at position
XX 619 and/or a substitution of A for G at position 898 of the OPTN coding
XX sequence. The mutation(s) is detected using a nucleic acid amplification
XX method using primers specific for the different exons of the coding
XX sequence, including primers SF10 ADL14975 and SR10 ADL14976 for exon 10.
XX
XX Sequence 1116 BP; 379 A; 172 C; 230 G; 335 T; 0 U; 0 Other;
XX
XX Query Match      100.0%; Score 22; DB 12; Length 1116;
XX Best Local Similarity 100.0%; Pred. No. 0.76;
XX Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCATGCTCACATTAACTGGA 22
   |||||
Db 786 TCATGCTCACATTAACTGGA 765

RESULT 4
ADW86578/c
ID ADW86578 standard; DNA; 1116 BP.
XX
XX ADW86578;
XX
XX 21-APR-2005 (first entry)
XX

```



```

XX DE Human optineurin (OPTN) gene exon 10 Seq 8.
XX KW glaucoma; optineurin; ophthalmological; ds; ocular disease;
XX KW DNA amplification; genetic marker.
XX OS Homo sapiens.
XX PN JP2005034112-A.
XX PD 10-FEB-2005.
XX PF 29-JUL-2003; 2003JP-00281897.
XX PR 02-AUG-2002; 2002JP-00226612.
XX PR 30-JUN-2003; 2003JP-00188070.
XX PA (TOAI-) TOA IYO DENSHI KK.
XX DR WPI; 2005-156038/17.
XX PT Estimating risk of onset of glaucoma, involves analyzing mutation in any
PT PT one portion of Optic new phosphorous gene, and utilizing analyzed
PT PT mutation as index for estimating risk of onset of glaucoma.
XX PS Example 1; SEQ ID NO 8; 13pp; Japanese.
XX CC This invention relates to a novel method for estimating the risk of onset
CC of glaucoma. Specifically, it refers to a method that involves analyzing
CC a mutation in any one portion of an optineurin (OPTN) gene and utilizing
CC the analyzed mutation as an index for estimating the risk of onset of
CC glaucoma. The present invention describes oligonucleotides to detect
CC mutations that hybridize with one or more portions of the OPTN glaucoma
CC related gene. Accordingly, it provides oligos that detect A619G and G898A
CC mutations in the human OPTN gene of patients at risk of developing
CC glaucoma, in particular primary open-angle glaucoma and/or normal tension
CC glaucoma. These mutations will not be present in a sample obtained from a
CC healthy person who is not at risk of onset of glaucoma. This
CC polynucleotide sequence is a human OPTN exon of the invention.
XX SQ Sequence 1116 BP; 379 A; 172 C; 230 G; 335 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 14; Length 1116;
Best Local Similarity 100.0%; Pred. No. 0.76;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCATGCTCACACATTAACTGGA 22
DB 786 TCATGCTCACACATTAACTGGA 765

RESULT 5
AD E13891/c
ID ADE13891 standard; DNA; 46951 BP.
XX AC ADE13891;
XX DT 29-JAN-2004 (first entry)
XX DE Human optineurin gene.
XX KW Human; optineurin; ds; gene; ophthalmological;
KW single nucleotide polymorphism; SNP; glaucoma;
KW progressive ocular hypertensive disorder; glaucoma related disorder.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT variation replace(391,G)
XX FT /tag= a
XX FT /standard name= "Single nucleotide polymorphism"
XX FT /tag= b
XX FT variation replace(691,G)
XX FT /tag= b

```

```

FT variation
FT /standard_name= "Single nucleotide polymorphism"
FT /tag= c
FT /standard_name= "Single nucleotide polymorphism"
FT /tag= d
FT /standard_name= "Single nucleotide polymorphism"
FT /tag= e
FT /standard_name= "Single nucleotide polymorphism"
FT /tag= f
FT /standard_name= "Single nucleotide polymorphism"
FT /tag= g
FT /standard_name= "Single nucleotide polymorphism"
FT /tag= h
FT /standard_name= "Single nucleotide polymorphism"
FT /tag= i
FT /standard_name= "Single nucleotide polymorphism"
FT /tag= j
FT /standard_name= "Single nucleotide polymorphism"
FT /tag= k
FT /standard_name= "Single nucleotide polymorphism"
FT /tag= l
FT /standard_name= "Single nucleotide polymorphism"
FT /tag= m
FT /standard_name= "Single nucleotide polymorphism"
FT /tag= n
FT /standard_name= "Single nucleotide polymorphism"
FT /tag= o
FT /standard_name= "Single nucleotide polymorphism"
FT /tag= p
FT /standard_name= "Single nucleotide polymorphism"
FT /tag= q
FT /standard_name= "Single nucleotide polymorphism"
FT /tag= r
FT /standard_name= "Single nucleotide polymorphism"
XX PN US2003190617-A1.
XX PD 09-OCT-2003.
XX PF 06-MAR-2002; 2002US-00091281.
XX PR 06-MAR-2002; 2002US-00091281.
XX PA (SIEE/) SI E.
XX PA (RAYM/) RAYMOND V.
XX PA (MORI/) MORISSETTE J.
XX PI Raymond V, Morissette J, Si E;
XX DR WPI; 2003-864168/80.
XX PT New nucleic acid sequences of the optineurin gene are useful to detect
XX PT polymorphisms particularly single nucleotide polymorphisms in the
XX PT optineurin promoter to diagnose, prognose and treat glaucoma and related
XX PS disorders.
XX PS Disclosure; SEQ ID NO 2; 159pp; English.
XX

```



PI Dmanac RT, Liu C, Tang YT;  
 XX WPI; 2001-639362/73.  
 DR P-PSDB; ABG28236.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 PT  
 PS Claim 1; SEQ ID NO 28227; 103pp; English.  
 PS  
 XX  
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activities. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
 CC coding sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 4557 BP; 1276 A; 1147 C; 1068 G; 1066 T; 0 U; 0 Other;  
 Query Match 88.2%; Score 19.4; DB 5; Length 4557;  
 Best Local Similarity 95.2%; Pred. No. 18;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 CATGCTCACACATTAAGTGA 22  
 Db 2891 CATGCTTACACATTAAGTGA 2911  
 RESULT 8  
 ADK15078  
 ID ADK15078 standard; DNA; 8277 BP.  
 AC ADK15078;  
 XX  
 XX 06-MAY-2004 (first entry)  
 DE Urinary specific nucleic acid #69.  
 XX  
 KW db; cytostatic; vaccine; gene therapy; urinary specific nucleic acid;  
 KW diagnosis; urinary cancer; transgenic animal.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003057839-A2.  
 XX  
 PD 17-JUL-2003.  
 XX  
 PF 23-DEC-2002; 2002WO-US041613.  
 XX  
 XX 28-DEC-2001; 2001US-0343690P.  
 XX  
 XX (DIAD-) DIADEXUS INC.  
 XX  
 PI Macina RA, Sun Y, Liu C;  
 PI WPI; 2003-587114/55.  
 DR

XX New urinary specific nucleic acid molecules and polypeptides, useful in  
 PT gene therapy, or for identifying, diagnosing, monitoring, staging,  
 PT imaging or treating urinary cancer and non-cancerous disease states of  
 PT the urinary tract.  
 XX  
 PS Claim 1; SEQ ID NO 69; 392pp; English.  
 PS  
 XX  
 CC The invention relates to novel urinary specific genes and their encoded  
 CC proteins or a nucleic acid molecule that selectively hybridizes, or has  
 CC at least 95 % sequence identity, to these nucleic acid molecules. The  
 CC urinary specific nucleic acid molecules, polypeptides and antibodies  
 CC against the polypeptides are useful for identifying, diagnosing,  
 CC monitoring, staging, imaging or treating urinary cancer and non-cancerous  
 CC disease states of the urinary tract. The nucleic acids and polypeptides  
 CC are also useful in gene therapy, for designing and/or identifying  
 CC antagonists or agonists of the polypeptides and for producing transgenic  
 CC animals and cells. This sequence corresponds to one of the urinary  
 CC specific nucleic acids of the invention.  
 XX  
 SQ Sequence 8277 BP; 2599 A; 1599 C; 1715 G; 2364 T; 0 U; 0 Other;  
 Query Match 88.2%; Score 19.4; DB 10; Length 8277;  
 Best Local Similarity 95.2%; Pred. No. 20;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 CATGCTCACACATTAAGTGA 22  
 Db 2746 CATGCTTACACATTAAGTGA 2766  
 RESULT 9  
 ADX61760  
 ID ADX61760 standard; cDNA; 842 BP.  
 XX  
 AC ADX61760;  
 XX  
 XX 21-APR-2005 (first entry)  
 DT  
 XX Plant full length insert polynucleotide seqid 32603.  
 DE  
 XX  
 KW plant protectant; plant growth regulant; gene therapy; plant;  
 KW recombinant DNA construct; physical array; plant breeding marker;  
 KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;  
 KW extreme osmotic condition; pathogen tolerance; pest tolerance;  
 KW growth rate; cell cycle pathway; disease resistance;  
 KW Galactomannan production; lignin production; plant growth regulator;  
 KW yield; plant growth; plant development; seed oil; protein yield;  
 KW protein content; gene; ss.  
 XX  
 OS Unidentified.  
 XX  
 XX US2004034888-A1.  
 PN  
 XX 19-FEB-2004.  
 PD  
 XX  
 PF 28-APR-2003; 2003US-00425114.  
 XX  
 PR 06-MAY-1999; 99US-00304517.  
 PR 05-NOV-2001; 2001US-00985678.  
 XX  
 XX (LIU/) LIU J.  
 PA (ZHOU/) ZHOU Y.  
 PA (KOVA/) KOVALIC D K.  
 PA (SCRE/) SCREEN S E.  
 PA (TABA/) TABASKA J E.  
 PA (CAOY/) CAO Y.  
 XX  
 PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JB, Cao Y;  
 XX WPI; 2004-180133/17.  
 DR  
 XX New recombinant DNA construct, useful for improving plant tolerance to

PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or  
PT pests, for conferring increased resistance to plant disease, or for  
PT improving yield.  
XX

PS Claim 1; SEQ ID NO 32603; 15pp; English.

XX The invention describes a recombinant DNA construct comprising a  
CC polynucleotide consisting of a sequence encoding an amino acid sequence  
CC available in electronic form from the US patent office at  
CC ftp.segdata.uspto.gov/sequence.html?docid:2004034888. The polynucleotide  
CC of the invention are also useful in physical arrays of molecules and as  
CC plant breeding markers. The recombinant DNA construct is useful for  
CC improving plant tolerance to cold, heat, drought, herbicides, extreme  
CC osmotic conditions, pathogens or pests, for manipulating growth rate in  
CC plant cells by modification of the cell cycle pathway, for conferring  
CC increased resistance to plant disease, for producing galactomannan,  
CC lignin or plant growth regulators, for increasing the rate of homologous  
CC recombination in plants, for improving yield by modification of  
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake  
CC or by providing improved plant growth and development under at least one  
CC stress condition or for modifying seed oil or protein yield and/or  
CC content. This sequence represents a plant full length insert  
CC polynucleotide that can be used in the recombinant DNA construct of the  
XX invention.

SQ Sequence 842 BP; 169 A; 325 C; 237 G; 111 T; 0 U; 0 Other;

Query Match 80.9%; Score 17.8; DB 13; Length 842;  
Best Local Similarity 90.5%; Pred. No. 92;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTAACTGGA 22

Db 49 CATGAGCACACATTAACTGGA 69

#### RESULT 10

ABD32968\_2/c  
Continuation (3 of 8) of ABD32968 from base 200001 (Human cancer-associated genomic DNA  
WP Sequence split into 8 fragments LOCUS ABD32968 Accession ABD32968

WP	Fragment Name	Begin	End
WP	ABD32968_1	1	110000
WP	ABD32968_2	100001	210000
WP	ABD32968_3	200001	310000
WP	ABD32968_4	300001	410000
WP	ABD32968_5	400001	510000
WP	ABD32968_6	500001	610000
WP	ABD32968_7	600001	710000
WP	ABD32968_8	700001	779603

Query Match 78.2%; Score 17.2; DB 13; Length 110000;  
Best Local Similarity 86.4%; Pred. No. 3.5e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TCATGCTCACACATTAACTGGA 22

Db 18693 TCATTACACACATTAACTGGA 18672

#### RESULT 11

ADZ59508  
ID ADZ59508 standard; DNA; 27251 BP.

XX ADZ59508;

XX 30-JUN-2005 (first entry)

XX Secondary hyperparathyroidism detection human polymorphic gene, CALCRL.

XX secondary hyperparathyroidism; endocrine-gen.; antithyroid;  
KW renal failure; nephrotropic; SNP detection;  
KW single nucleotide polymorphism; SNP; gene; ds; CALCRL.

XX

OS Homo sapiens.  
XX Key Location/Qualifiers  
FH variation replace(17251,G)  
FT /\*tag= a  
FT /standard\_name= "single nucleotide polymorphism"

XX JP2005102601-A.

XX 21-APR-2005.

XX 30-SEP-2003; 2003JP-00341015.

XX 30-SEP-2003; 2003JP-00341015.

XX (HYUB-) HYUBITTO GENOMICS KK.  
XX (JIKE-) UNIV JIKEI.

XX WPI; 2005-358641/37.

XX Testing secondary hyperparathyroidism in chronic renal failure patient.  
PT involves detecting variation in gene chosen from CACNA1C, CALCRL, CH13L1,  
PT EGF, FGF1, GFR1, GPR56 and GPRK6.

XX Claim 4; SEQ ID NO 2; 138pp; Japanese.

XX The invention relates to a novel method for testing secondary  
CC hyperparathyroidism in a chronic renal failure patient. The method  
CC involves detecting a variation in a gene chosen from CACNA1C, CALCRL,  
CC CH13L1, EGF, FGF1, GPR56, GPRK6, IL10RB, IL12RB1, KCNJ14,  
CC KCNQ1, ORCTL4, PDGFRA, SCY12A1, SLC2A3, TGFBR3, TMEM1, CALCR,  
CC IL17R, OSTF1, FGF6, HGF, MET, TGFBI and VEGF, or detecting the base in a  
CC polymorphism region existing in the vicinity of any one of the genes. The  
CC invention further comprises a reagent or kit for testing secondary  
CC hyperparathyroidism in a chronic renal failure patient. This  
CC polynucleotide sequence represents the polymorphism containing human  
CC CALCRL gene of the invention.

SQ Sequence 27251 BP; 9618 A; 4481 C; 4480 G; 8672 T; 0 U; 0 Other;

Query Match 76.4%; Score 16.8; DB 14; Length 27251;  
Best Local Similarity 90.0%; Pred. No. 4.6e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCATGCTCACACATTAACTG 20

Db 16362 TCATGCATCTCACATTAACTG 16381

#### RESULT 12

ACN44934\_2/c  
Continuation (3 of 4) of ACN44934 from base 200001 (Human genomic sequence HCG1811271.  
WP Sequence split into 4 fragments LOCUS ACN44934 Accession ACN44934

WP	Fragment Name	Begin	End
WP	ACN44934_0	1	110000
WP	ACN44934_1	100001	210000
WP	ACN44934_2	200001	310000
WP	ACN44934_3	300001	401616

Query Match 76.4%; Score 16.8; DB 11; Length 110000;  
Best Local Similarity 90.0%; Pred. No. 5.6e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 ATGCTCACACATTAACTGGA 22

Db 69113 ATGCTCCACACTAACTGGA 69094

#### RESULT 13

ADZ46976\_4/c

Continuation (5 of 6) of ADZ46976 from base 400001 (Human phosphodiesterase 4B DNA SEQ  
WP Sequence split into 6 fragments LOCUS ADZ46976 Accession ADZ46976  
WP Fragment Name Begin End



sequence; (2) a transformed host cell comprising the recombinant construct; (3) a method to isolate (1) encoding ryanodine receptors and related polypeptides; (4) an isolated polypeptide having ryanodine receptor activity; (5) a method for evaluating at least one compound for its ability to modulate calcium homeostasis; (6) a method for evaluating at least one compound which modulates ryanodine receptor activity; (7) an isolated nucleic acid fragment encoding an insect ion channel comprising at least two fully defined polypeptide sequences selected from SEQ ID NOS. 63-119 provided that the polypeptide sequences is not SEQ ID NO. 56, 120-126; (8) a method for identifying a nucleic acid sequence encoding an insect ion channel; (9) a method for expressing an isolated nucleic acid fragment encoding a toxic insect ion channel; and (10) recombinant construct comprising in the 5' to 3' direction a promoter operably linked to an isolated nucleic acid fragment encoding a toxic insect ion channel. The isolated nucleotide fragment (1) encoding a ryanodine receptor is useful for the isolation of other pest ryanodine receptors and developing of screens to identify insecticidally active compounds. The nucleic acid fragments are useful as pesticides, fragments of protein for antibody production, fragments of protein for determination of the structure of insecticide binding sites and in the identification of insecticides that disrupt the calcium balance in cells through other messengers that interact with the receptor calcium release mechanism. The present sequence is used in the exemplification of the present invention.

Sequence 15845 BP; 5273 A; 2716 C; 3325 G; 4528 T; 0 U; 3 Other;

Query Match 74.5%; Score 16.4; DB 12; Length 15845;  
Best Local Similarity 94.4%; Pred. No. 6.8e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTAACT 19

Db 5599 CATGCTCACACATTGACT 5616

RESULT 16

ADQ59419/c  
ID ADQ59419 standard; DNA; 247544 BP.

AC ADQ59419;

XX 07-OCT-2004 (first entry)

XX Human cancer-associated (CA) gene sequence SEQ ID NO:55.

XX human; cancer-associated gene; cancer-associated protein; cytostatic;

KW gene therapy; vaccine; tyrosine kinase antagonist;

KW G-protein coupled receptor antagonist; cancer; lymphoma; gene; ds.

XX Homo sapiens.

XX WO2004058288-A1.

XX 15-JUL-2004.

XX 15-DEC-2003; 2003WO-US040082.

XX 17-DEC-2002; 2002US-00322696.

XX (SAGR-) SAGRES DISCOVERY INC.

XX Morris DW, Malandro MS;

XX WPI; 2004-543349/52.

XX P-PSDB; ADQ59421.

XX New cancer-associated nucleic acid for diagnosing, preventing or treating cancer (e.g. lymphoma) or for screening agents that may be used for treating or preventing cancer.

XX Disclosure; SEQ ID NO 55; 143pp; English.

XX The present invention describes human cancer-associated (CA) nucleotide

sequences (1). Also described: (1) an expression vector comprising (1); (2) a host cell comprising (1) or the expression vector; (3) a microarray for detecting a CA nucleic acid; (4) an isolated polypeptide encoded within an open reading frame of a CA sequence; (5) an isolated antibody, or its antigen binding fragment, that binds to the above polypeptide; (6) a hybridoma that produces the monoclonal antibody described above; (7) a pharmaceutical composition comprising the antibody and a pharmaceutical excipient; (8) a kit for detecting or diagnosing cancer cells, comprising the above (monoclonal) antibody or polynucleotide that selectively hybridizes to any of the polynucleotide sequences mentioned above; (9) methods for diagnosing cancer or for detecting the presence or absence of cancer cells in an individual; (10) a method for inhibiting growth of cancer cells in an individual; (11) a method for delivering a therapeutic agent to cancer cells in an individual; (12) an electronic library comprising the polynucleotide or polypeptide, or their fragments, mentioned above; (13) a method of screening for anticancer activity; (14) methods for detecting cancer associated with expression of a polypeptide or the presence of the antibody in a test cell or serum sample; (15) a method for screening for a bioactive agent capable of modulating the activity of a CA protein encoded by the above nucleic acid molecule; and (16) a method for treating cancers. (1) has cytostatic activity, and can be used in gene therapy, in vaccines, as a tyrosine kinase antagonist, and as a G-protein coupled receptor antagonist. The compositions and methods of the present invention can be used for diagnosing, preventing and treating cancer, especially lymphomas. They may also be used in screening for agents that may be used for treating or preventing cancer. The present sequence represents a human CA gene sequence, which is given in the exemplification of the present invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 247544 BP; 60925 A; 55098 C; 58499 G; 69875 T; 0 U; 3147 Other;

Query Match 74.5%; Score 16.4; DB 12; Length 247544;

Best Local Similarity 94.4%; Pred. No. 9.7e+02;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 TGCTCACACATTAACTGG 21

Db 117312 TGCTCACACATTCACTGG 117295

RESULT 17

ADZ13712/c

ID ADZ13712 standard; DNA; 247654 BP.

AC ADZ13712;

XX 16-JUN-2005 (first entry)

XX Murine cancer-associated genomic DNA #107.

XX Diagnosis; DNA microarray; microarray; biochip; cancer; neoplasm;

KW cytostatic; gene; ds.

XX Mus sp.

XX WO2005031001-A2.

XX 07-APR-2005.

XX 23-SEP-2004; 2004WO-US031617.

XX 23-SEP-2003; 2003US-00669920.

XX (CHIR ) CHIRON CORP.

XX Morris DW, Malandro MS;

XX WPI; 2005-273395/28.

XX Nucleic acid array useful for detecting cancer associated nucleic acid,

PT comprises two or more nucleic acid probes.  
 XX Disclosure; SEQ ID NO 1232; 198pp; English.  
 XX  
 CC The invention relates to a nucleic acid array for detecting a cancer  
 CC associated (CA) nucleic acid, comprising two or more nucleic acid probes.  
 CC The invention also relates to a peptide array comprising two or more  
 CC isolated polypeptides encoded by a CA nucleic acid sequence, a compound  
 CC that binds to a polypeptide, an isolated antibody or its fragment which  
 CC binds to a polypeptide, which is prepared by immunizing a host animal  
 CC with a composition comprising the polypeptide or its antigen binding  
 CC fragment and collecting cells from the host expressing antibodies against  
 CC the antigen or its antigen binding fragment, a composition comprising the  
 CC antibody and a carrier, a method of screening for anticancer activity, a  
 CC method of detecting a CA nucleic acid, a method of diagnosing cancer, a  
 CC method of treating cancer and a method of inhibiting expression of a CA  
 CC nucleic acid in a cell. The CA nucleic acids are useful for detecting CA  
 CC nucleic acids. The antibody is useful for detecting the presence or  
 CC absence of cancer cells in an individual which involves contacting cells  
 CC from the individual with the antibody and detecting a complex of a CA  
 CC protein from the cancer cells and the antibody, where the detection of  
 CC the complex correlates with the presence of cancer cells in the  
 CC individual. The composition is useful for inhibiting growth of cancer  
 CC cells in an individual or for delivering a therapeutic agent to cancer  
 CC cells in an individual. The invention is also useful for diagnosing  
 CC cancer, for treating cancer and for inhibiting expression of a CA gene in  
 CC a cell. This sequence represents murine cancer-associated genomic DNA of  
 CC the invention.  
 XX  
 SQ Sequence 247654 BP; 60955 A; 55129 C; 58524 G; 69899 T; 0 U; 3147 Other;

Query Match 74.5%; Score 16.4; DB 14; Length 247654;  
 Best Local Similarity 94.4%; Pred. No. 9.7e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 TGCTCACACATTCACTGG 21

DB 117422 TGCTCACACATTCACTGG 117405

RESULT 18  
 ADG89121  
 ID ADG89121 standard; DNA; 65 BP.  
 XX  
 AC ADG89121;  
 XX  
 DT 11-MAR-2004 (first entry)  
 XX  
 DE Cancer detection method related oligonucleotide #69.  
 XX  
 KW ss; cancer; gene expression;  
 XX estrogen receptor-positive invasive breast cancer.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO2003078662-A1.  
 XX  
 XX 25-SEP-2003.  
 XX  
 XX 12-MAR-2003; 2003WO-US0007713.  
 XX  
 XX 13-MAR-2002; 2002US-0364890P.  
 XX  
 XX 18-SEP-2002; 2002US-0412049P.  
 XX  
 XX (GENO-) GENOMIC HEALTH INC.  
 XX  
 XX Baker JB, Cronin MT, Kiefer MC, Shak S, Walker MG;  
 XX WPI; 2003-767536/72.  
 XX

PT Predicting clinical outcome for a patient diagnosed with cancer comprises  
 PT determining the expression level of one or more genes, and compared to  
 PT the amount found in a reference cancer tissue set.

XX

PS Disclosure; SEQ ID NO 69; 198pp; English.

XX  
 CC The invention relates to a method of predicting clinical outcome for a  
 CC patient diagnosed with cancer by determining the expression level of one  
 CC or more genes, or their expression products, selected from p53BP2,  
 CC cathepsin B, cathepsin L, Ki67/MiB1, and thymidine kinase in a cancer  
 CC tissue obtained from the patient, normalized against control gene(s), and  
 CC compared to the amount found in a reference cancer tissue set. The  
 CC specification also discloses an array comprising polynucleotides  
 CC hybridizing to the following genes: FOXM1, PRAME, Bcl2, STK15, CRGPI, Ki-  
 CC 67, GSTM1, CA9, PR, BCC3, NME1, SURV, CAT3A3, TP53, YB-1, BPIYD, GSTM3,  
 CC RPS6KB1, Sro, Chk1, ID1, ESR1, p27, CCN1, XIAP, Chk2, CDC25B, IGFBP3,  
 CC AKO55699, PI3KC2A, TGFBR3, BAG1, CYP3A4, EPCAM, VEGFC, PS2, HENT1, WFS1,  
 CC HNF3A, NFKBp65, BRCA2, EGFR, TK1, VDR, Contig51037, PENT1, EPHX1, IFTA,  
 CC CDH1, HIF1T, IGFBP3, CTSE, Her2 and DIABLO, immobilized on a solid  
 CC surface. The methods are useful for predicting clinical outcome for a  
 CC patient diagnosed with cancer, classifying cancer, and predicting the  
 CC likelihood of long-term survival of a breast cancer patient, or a patient  
 CC diagnosed with invasive breast cancer or with estrogen receptor (ER)-  
 CC positive invasive breast cancer. This sequence corresponds to an  
 CC oligonucleotide used in the method of the invention.  
 XX

SQ Sequence 65 BP; 17 A; 17 C; 17 G; 14 T; 0 U; 0 Other;

Query Match 73.6%; Score 16.2; DB 10; Length 65;  
 Best Local Similarity 85.7%; Pred. No. 4.2e+02;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTCACTGGA 22

DB 41 CATGCCACACATCACTGGA 61

RESULT 19

AD112599  
 ID AD112599 standard; cDNA; 210 BP.

XX  
 AC AD112599;

XX 22-APR-2004 (first entry)

XX DE Mutant human BRCA1 cDNA resulting from frameshift of deletion 6 Seq82.

XX ss; cancer; human; tumour suppressor;  
 KW breast cancer susceptibility gene 1; BRCA1; repetitive Alu;  
 KW ovarian cancer; recombination; mutant.

XX OS Homo sapiens.

XX WO2003104474-A2.

XX 18-DEC-2003.

XX 09-JUN-2003; 2003WO-US018098.

XX 07-JUN-2002; 2002US-0387132P.

XX 09-AUG-2002; 2002US-0402430P.

XX (MYRI-) MYRIAD GENETICS INC.

XX Scholl T, Hendrickson BC, Ward B, Pruss D;

XX WPI; 2004-062369/06.

XX Predicting a predisposition to cancer in a patient comprising detecting a  
 PT deletion in the BRCA1 gene that results from the unequal crossover  
 PT between a pair of repetitive sequences in the BRCA1 gene.

XX Disclosure; SEQ ID NO 82; 59pp; English.

XX This invention relates to a novel method for predicting a predisposition  
 CC to cancer in a patient by detecting large deletions in the human tumour



CC suppressor gene identified as the breast cancer susceptibility gene 1  
 CC (BRCA1). Specifically, it refers to deletions that result from the  
 CC unequal crossover between a pair of repetitive Alu sequences in the BRCA1  
 CC gene, such that the recombined nucleotide sequence containing the  
 CC deletion indicates a predisposition to breast and ovarian cancer. The  
 CC present invention describes newly discovered deletion mutations that are  
 CC believed to be deleterious and cause significant alterations in the  
 CC structure or biochemical function of BRCA1. Accordingly, it provides  
 CC methods for detecting such mutants, as well as identifying and screening  
 CC for cytostatic compounds useful for treating or preventing cancers  
 CC associated with a BRCA1 genetic variant. This polynucleotide is a mutant  
 CC human BRCA1 cDNA fragment that arises from a frame shift mutation due to  
 CC a recombination event (deletion 6), which causes the omission of exons 14  
 CC -20, given in an exemplification of the invention.

XX Sequence 210 BP; 47 A; 48 C; 61 G; 54 T; 0 U; 0 Other;

Query Match 73.6%; Score 16.2; DB 12; Length 210;  
 Best Local Similarity 85.7%; Pred. No. 4.9e+02;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTAACTGGA 22  
 Db 46 CATGCCACAGATCAACTGGA 66  
 ||||| ||||| ||||| |||||

RESULT 20  
 AD112627  
 ID AD112627 standard; DNA; 390 BP.  
 XX  
 AC AD112627;  
 XX  
 DT 22-APR-2004 (first entry)  
 XX  
 DE Human BRCA1 DNA fragment comprising deletion 6.  
 XX  
 KW ds; cancer; human; tumour suppressor;  
 KW breast cancer susceptibility gene 1; BRCA1; repetitive Alu;  
 KW ovarian cancer; recombination; mutant; gene.

XX Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS  
 FT 1..390  
 FT /\*tag= a  
 FT /partial  
 FT /product= "BRCA1 mutant protein"  
 FT /note= "Start codon is absent"

XX WO2003104474-A2.  
 XX  
 XX 18-DEC-2003.  
 XX  
 XX 09-JUN-2003; 2003WO-US018098.  
 XX  
 XX 07-JUN-2002; 2002US-0387132P.  
 XX 09-AUG-2002; 2002US-0402430P.  
 XX  
 XX (MYRI-) MYRIAD GENETICS INC.

XX Scholl T, Hendrickson BC, Ward B, Pruss D;  
 XX WPI; 2004-062369/06.  
 XX P-PSDB; AD112628.

XX Predicting a predisposition to cancer in a patient comprising detecting a  
 XX deletion in the BRCA1 gene that results from the unequal crossover  
 XX between a pair of repetitive sequences in the BRCA1 gene.  
 XX Disclosure; Fig 9; 59pp; English.  
 XX  
 XX This invention relates to a novel method for predicting a predisposition  
 XX to cancer in a patient by detecting large deletions in the human tumour

CC suppressor gene identified as the breast cancer susceptibility gene 1  
 CC (BRCA1). Specifically, it refers to deletions that result from the  
 CC unequal crossover between a pair of repetitive Alu sequences in the BRCA1  
 CC gene, such that the recombined nucleotide sequence containing the  
 CC deletion indicates a predisposition to breast and ovarian cancer. The  
 CC present invention describes newly discovered deletion mutations that are  
 CC believed to be deleterious and cause significant alterations in the  
 CC structure or biochemical function of BRCA1. Accordingly, it provides  
 CC methods for detecting such mutants, as well as identifying and screening  
 CC for cytostatic compounds useful for treating or preventing cancers  
 CC associated with a BRCA1 genetic variant. This polynucleotide is the human  
 CC BRCA1 DNA mutant that comprises the recombination event (deletion 6) that  
 CC causes the omission of exons 14-20, given in an exemplification of the  
 CC invention.

XX Sequence 390 BP; 109 A; 96 C; 98 G; 87 T; 0 U; 0 Other;

Query Match 73.6%; Score 16.2; DB 12; Length 390;  
 Best Local Similarity 85.7%; Pred. No. 5.3e+02;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTAACTGGA 22  
 Db 226 CATGCCACAGATCAACTGGA 246  
 ||||| ||||| ||||| |||||

RESULT 21  
 ACH48346/c  
 ID ACH48346 standard; cDNA; 456 BP.  
 XX  
 AC ACH48346;

DT 13-OCT-2003 (first entry)

XX Human lung tumour cDNA #479.

XX Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;  
 KW genome mapping; biodiversity; genetic disorder.

XX Homo sapiens.

XX US2003073623-A1.

XX 17-APR-2003.

XX 30-JUL-2001; 2001US-00918995.

XX 30-JUL-2001; 2001US-00918995.

XX (DRNA/) DRMANAC R T.

XX (LABA/) LABAT I.

XX (STAC/) STACHE-CRAIN B.

XX (DICK/) DICKSON M C.

XX (JONE/) JONES L W.

XX Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;

XX WPI; 2003-615964/58.

XX New polynucleotide sequences obtained from various cDNA libraries, useful  
 XX as hybridization probes, as oligomers for PCR, for chromosome and gene  
 XX mapping, in the recombinant production of protein, or in generating  
 XX antisense DNA or RNA.

XX Claim 1; SEQ ID NO 35558; 44pp; English.

XX The invention relates to an isolated polynucleotide comprising any one of  
 XX 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was  
 XX determined by the technique of SBH (sequencing by hybridisation). Also  
 XX included is a purified polypeptide comprising a sequence corresponding to  
 XX a reading frame of the novel polynucleotide. The nucleic acid sequences  
 XX are useful in diagnostics as expressed sequence tags (EST) for  
 XX identifying expressed genes or for physical mapping of the human genome,



CC in forensics, in assessing biodiversities, or in identifying mutations  
 CC responsible for genetic disorders and other traits. The nucleotide  
 CC sequences are also useful as hybridisation probes, as oligomers for PCR,  
 CC for chromosome and gene mapping, in the recombinant production of  
 CC protein, or in generating antisense DNA or RNA. The purified polypeptide  
 CC is useful for generating antibodies specific for it. The present sequence  
 CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data  
 CC for this patent did not form part of the printed specification, but was  
 CC obtained in electronic format directly from USPTO at  
 CC seqdata.uspto.gov/sequence.html?DocID=20030073623

XX SQ Sequence 456 BP; 137 A; 85 C; 109 G; 125 T; 0 U; 0 Other;

Query Match 73.6%; Score 16.2; DB 9; Length 456;

Best Local Similarity 85.7%; Pred. No. 5.4e+02;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTAACTGGA 22

||||| ||||| ||||| |||||

Db 301 CATGCTGACACATTTTCTGGA 281

RESULT 22

AA009872

ID AAD09872 standard; cDNA; 1083 BP.

XX AC AAD09872;

XX DT 11-SEP-2003 (revised)

XX DT 18-SEP-2001 (first entry)

XX DE Infectious salmon anaemia virus (ISAV) 92-T cDNA.

XX KW Infectious salmon anaemia virus; ISAV; 92-T clone; vaccine; prophylaxis;

XX KW Infectious salmon anaemia; therapy; antianaemic; ss.

XX OS Infectious salmon anemia virus.

XX FH Key Location/Qualifiers

XX CDS 1. .1083

XX FT /tag= a

XX FT /product= "ISAV 92-T protein"

XX PN WO200149712-A2.

XX PD 12-JUL-2001.

XX PF 03-JAN-2001; 2001WO-EP0000046.

XX PR 07-JAN-2000; 2000EP-00200054.

XX PR 29-FEB-2000; 2000EP-00200700.

XX PA (ALKU ) AKZO NOBEL NV.

XX PI Biering E, Krossoy B;

XX WPI; 2001-441845/47.

XX DR P-PSDB; AAE05125.

XX PT Novel vaccine for treatment and/or prevention of infectious salmon anemia

XX PT in fish.

XX PS Claim 5; Page 23-25; 43pp; English.

XX CC The present invention relates to vaccine for prevention and/or

XX CC prophylaxis of infectious salmon anaemia in fish. The invention provides

XX CC for nucleic acid sequence encoding viral proteins of infectious salmon

XX CC anaemia virus (ISAV) as well as the isolated protein. Nucleic acids

XX CC encoding viral protein is useful for the manufacture of a DNA vaccine for

XX CC diagnosis, treatment and/or prophylaxis of infectious salmon anaemia in

XX CC fish, and viral protein can be used for the manufacture of antibodies

XX CC that are specific for ISAV. The present sequence is infectious salmon

XX CC anaemia virus 92-T cDNA. (Updated on 11-SEP-2003 to standardise OS field)

XX

SQ Sequence 1083 BP; 300 A; 211 C; 288 G; 284 T; 0 U; 0 Other;

Query Match 73.6%; Score 16.2; DB 4; Length 1083;

Best Local Similarity 85.7%; Pred. No. 6e+02;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTAACTGGA 22

||||| ||||| ||||| |||||

Db 749 CATGCCACACATTGACTGGA 769

RESULT 23

AA009875

ID AAD09875 standard; cDNA; 1166 BP.

XX AC AAD09875;

XX DT 18-SEP-2001 (first entry)

XX DE Infectious salmon anaemia virus (ISAV) 92-M fusion cDNA.

XX KW Infectious salmon anaemia virus; ISAV; 92-M clone; vaccine; prophylaxis;

XX KW Infectious salmon anaemia; therapy; fusion protein; antianaemic; ss.

XX OS Infectious salmon anemia virus.

XX OS Unidentified.

XX OS Chimeric.

XX FH Key Location/Qualifiers

XX CDS 6. .1166

XX FT /tag= a

XX FT /product= "ISAV 92-M protein"

XX FT sig\_peptide

XX FT 6. .68

XX FT /tag= b

XX FT /note= "Mellitin signal peptide"

XX FT misc\_feature

XX FT 69. .83

XX FT /tag= c

XX FT /note= "Linker sequence"

XX FT mat\_peptide

XX FT 84. .1163

XX FT /tag= d

XX FT /product= "Mature ISAV 92-M protein"

XX PN WO200149712-A2.

XX PD 12-JUL-2001.

XX PF 03-JAN-2001; 2001WO-EP0000046.

XX PR 07-JAN-2000; 2000EP-00200054.

XX PR 29-FEB-2000; 2000EP-00200700.

XX PA (ALKU ) AKZO NOBEL NV.

XX PI Biering E, Krossoy B;

XX WPI; 2001-441845/47.

XX DR P-PSDB; AAE05128.

XX PT Novel vaccine for treatment and/or prevention of infectious salmon anemia

XX PT in fish.

XX PS Claim 2; Page 37-39; 43pp; English.

XX CC The present invention relates to vaccine for prevention and/or

XX CC prophylaxis of infectious salmon anaemia in fish. The invention provides

XX CC for nucleic acid sequence encoding viral proteins of infectious salmon

XX CC anaemia virus (ISAV) as well as the isolated protein. Nucleic acids

XX CC encoding viral protein is useful for the manufacture of a DNA vaccine for

XX CC diagnosis, treatment and/or prophylaxis of infectious salmon anaemia in

XX CC fish, and viral protein can be used for the manufacture of antibodies

XX CC that are specific for ISAV. The present sequence is infectious salmon

XX CC anaemia virus 92-M fusion cDNA

XX	SQ	Sequence 1166 BP; 321 A; 228 C; 302 G; 315 T; 0 U; 0 Other;	
		Query Match 73.6%; Score 16.2; DB 4; Length 1166;	
		Best Local Similarity 85.7%; Pred. No. 6.1e+02;	
		Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;	
QY	2	CATGCTCACACATTAAGTGA 22	
DB	832	CATGCCAGACATTGACTGA 852	
RESULT 24			
AAD09874			
ID	AAD09874	standard; cDNA; 1167 BP.	
XX	AC	AAD09874;	
XX	DT	11-SEP-2003 (revised)	
DT	18-SEP-2001	(first entry)	
XX	XX	Infectious salmon anaemia virus (ISAV) 9Z cDNA.	
DE	XX	Infectious salmon anaemia virus; ISAV; 9Z clone; vaccine; prophylaxis;	
KW	XX	Infectious salmon anaemia; therapy; antianaemic; ss.	
OS	XX	Infectious salmon anaemia virus.	
XX	Key	Location/Qualifiers	
FT	CDS	1..1167	
FT		/*tag= a	
FT		/product= "ISAV 9Z protein"	
XX	WO200149712-A2.		
XX	12-JUL-2001.		
XX	03-JAN-2001; 2001WO-EP000046.		
XX	07-JAN-2000; 2000EP-00200054.		
PR	29-FEB-2000; 2000EP-00200700.		
XX	(ALKU ) AKZO NOBEL NV.		
XX	Biering E, Krossoy B;		
XX	WPI; 2001-441845/47.		
DR	P-PSDB; AAE05127.		
XX	Novel vaccine for treatment and/or prevention of infectious salmon anemia		
PT	in fish.		
XX	Claim 2; Page 33-35; 43pp; English.		
XX	The present invention relates to vaccine for prevention and/or		
CC	prophylaxis of infectious salmon anaemia in fish. The invention provides		
CC	for nucleic acid sequence encoding viral proteins of infectious salmon		
CC	anaemia virus (ISAV) as well as the isolated protein. Nucleic acids		
CC	encoding viral protein is useful for the manufacture of a DNA vaccine for		
CC	diagnosis, treatment and/or prophylaxis of infectious salmon anaemia in		
CC	fish, and viral protein can be used for the manufacture of antibodies		
CC	that are specific for ISAV. The present sequence is infectious salmon		
CC	anaemia virus 9Z cDNA. (Updated on 11-SEP-2003 to standardise OS field)		
XX	SQ	Sequence 1167 BP; 319 A; 233 C; 303 G; 312 T; 0 U; 0 Other;	
		Query Match 73.6%; Score 16.2; DB 4; Length 1167;	
		Best Local Similarity 85.7%; Pred. No. 6.1e+02;	
		Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;	
QY	2	CATGCTCACACATTAAGTGA 22	
DB	833	CATGCCAGACATTGACTGA 853	

RESULT 25			
ACA21535			
ID	ACA21535	standard; DNA; 1272 BP.	
XX	AC	ACA21535;	
XX	DT	19-JUN-2003 (first entry)	
XX	DE	Prokaryotic essential gene #3192.	
XX	KW	Antisense; ds; prokaryotic essential gene; cell proliferation;	
KW	XX	drug design; gene.	
XX	OS	Bacillus anthracis.	
XX	PN	WO200277183-A2.	
XX	PD	03-OCT-2002.	
XX	PF	21-MAR-2002; 2002WO-US009107.	
XX	PR	21-MAR-2001; 2001US-00815242.	
PR	06-SEP-2001; 2001US-00948993.		
PR	25-OCT-2001; 2001US-0342923P.		
PR	08-FEB-2002; 2002US-00072851.		
PR	06-MAR-2002; 2002US-0362699P.		
XX	(ELIT- ) ELITRA PHARM INC.		
XX	Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KU, Zvekind JW;		
PI	Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;		
XX	WPI; 2003-029926/02.		
DR	P-PSDB; ABU17665.		
XX	New antisense nucleic acids, useful for identifying proteins or screening		
PT	for homologous nucleic acids required for cellular proliferation to		
FT	isolate candidate molecules for rational drug discovery programs.		
XX	Claim 14; SEQ ID NO 9405; 1766pp; English.		
XX	The invention relates to an isolated nucleic acid comprising any one of		
CC	the 6213 antisense sequences given in the specification where expression		
CC	of the nucleic acid inhibits proliferation of a cell. Also included are:		
CC	(1) a vector comprising a promoter operably linked to the nucleic acid		
CC	encoding a polypeptide whose expression is inhibited by the antisense		
CC	nucleic acid; (2) a host cell containing the vector; (3) an isolated		
CC	polypeptide or its fragment whose expression is inhibited by the		
CC	antisense nucleic acid; (4) an antibody capable of specifically binding		
CC	the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular		
CC	proliferation or the activity of a gene in an operon required for		
CC	proliferation; (7) identifying a compound that influences the activity of		
CC	the gene product or that has an activity against a biological pathway		
CC	required for proliferation, or that inhibits cellular proliferation; (8)		
CC	identifying a gene required for cellular proliferation or the biological		
CC	pathway in which a proliferation-required gene or its gene product lies		
CC	or a gene on which the test compound that inhibits proliferation of an		
CC	organism acts; (9) manufacturing an antibiotic; (10) profiling a		
CC	compound's activity; (11) a culture comprising strains in which the gene		
CC	product is overexpressed or underexpressed; (12) determining the extent		
CC	to which each of the strains is present in a culture or collection of		
CC	strains; or (13) identifying the target of a compound that inhibits the		
CC	proliferation of an organism. The antisense nucleic acids are useful for		
CC	identifying proteins or screening for homologous nucleic acids required		
CC	for cellular proliferation to isolate candidate molecules for rational		
CC	drug discovery programs, or for screening homologous nucleic acids		
CC	required for proliferation in cells other than S. aureus, S. typhimurium,		
CC	K. pneumoniae or P. aeruginosa. The present sequence is one of the target		
CC	prokaryotic essential genes. Note: The sequence data for this patent did		
CC	not form part of the printed specification, but was obtained in		
CC	electronic format directly from WIPO at		

```

CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 1272 BP; 435 A; 209 C; 279 G; 349 T; 0 U; 0 Other;
Query Match 73.6%; Score 16.2; DB 8; Length 1272;
Best Local Similarity 85.7%; Pred. No. 6.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTAAGTGA 22
DB 1132 CATGTCATACATTAAATGGA 1152

RESULT 26
AAC88168/c
ID AAC88168 standard; RNA; 1320 BP.
XX AC AAC88168;
XX DT 14-MAR-2001 (first entry)
XX DE Infectious salmon anemia (ISA) virus gene segment 7 SEQ ID NO:1.
XX KW Infectious salmon anemia virus; ISA virus; gene segment 7; vaccine;
XX KW immunogenic; detection; diagnosis; antiviral; fish; aquatic organism;
XX KW prophylactic health care; salmon; ss.
XX OS Infectious salmon anemia virus.
XX PN WO200072878-A1.
XX PD 07-DEC-2000.
XX PF 29-MAY-2000; 2000WO-NO000179.
XX PR 31-MAY-1999; 99NO-00002608.
XX PA (GENO-) GENOMAR AS.
XX PI Rimstad E;
XX WPI; 2001-061447/07.
XX P-PSDB; AAB36676.
XX DNA which encodes protein from infectious salmon anemia virus, useful as
a vaccine against the virus in salmon, in prophylactic health care in
fish and aquatic organisms, in diagnostic systems and in biomedicine.
XX Claim 1; Page 19; 31pp; English.
XX The present invention describes a DNA (I) having a fully defined sequence
(AAC88168) of 1320 base pairs (bp) or sequences with at least 80%
homology to (S) and encoding a protein from infectious salmon anemia
(ISA) virus. Also described are: (1) a vector (II) containing (I); (2) a
vaccine (III) against ISA virus including (I); and (3) a diagnostic kit
comprising primers reacting with ISA virus sequences, for the detection
of ISA specific nucleic acids or proteins. (I) is useful for vaccine
purposes, prophylactic health care in fish and aquatic organisms, in
biomedicine (for example as model organisms for influenza-like virus), in
human medicine, in research and in diagnostic systems for the detection
of ISA virus specific protein/nucleic acids, and detection of antibodies
directed against the proteins. (I) is used within preventive medicine in
fish and (III) is used to protect salmon against ISA diseases. The
present sequence represents the ISA virus gene segment 7 DNA sequence,
which is antisense to the RNA sequence given in AAC88168
XX SQ Sequence 1320 BP; 364 A; 332 C; 249 G; 0 T; 375 U; 0 Other;
Query Match 73.6%; Score 16.2; DB 4; Length 1320;
Best Local Similarity 85.7%; Pred. No. 6.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTAAGTGA 22
DB 482 CATGCCACACATTGACTGGA 462

RESULT 28
AAC88168/c
ID ADK45527 standard; DNA; 2325 BP.
XX AC ADK45527;
XX DT 24-FEB-2005 (first entry)

```

XX Streptococcus pneumoniae gene, Seq ID No 2042.  
DE db; gene; Antibacterial; Gene therapy; Vaccine; Streptococcus pneumoniae.  
XX Streptococcus pneumoniae.  
XX US6699703-B1.  
XX OS  
XX PN  
XX US6699703-B1.  
XX PD  
XX 02-MAR-2004.  
XX XX  
XX 26-MAY-2000; 2000US-00583110.  
XX XX  
XX 02-JUL-1997; 97US-0051553P.  
XX PR 12-MAY-1998; 98US-0085131P.  
XX PR 30-JUN-1998; 98US-00107433.  
XX XX  
XX (GENO-) GENOME THERAPEUTICS CORP.  
XX PA Doucette-Stamm L, Bush D, Zeng Q, Opperman T, Housewartz CE;  
XX PI  
XX XX  
XX DR WPI; 2004-212399/20.  
XX DR P-PSDB; ADK48188.  
XX XX  
XX New nucleic acid molecules and polypeptides useful for diagnosing,  
PT preventing and treating pathological conditions resulting from bacterial  
PT infection, e.g. Streptococcus pneumoniae infection, and in drug  
PT screening.  
XX XX  
XX PS Disclosure; SEQ ID NO 2042; 301pp; English.  
XX XX  
XX The invention relates to isolated Streptococcus pneumoniae nucleic acids  
CC and polypeptides. The nucleic acids and proteins are useful for  
CC diagnosing, preventing and treating pathological conditions resulting  
CC from bacterial infection, such as S. pneumoniae infection. These may also  
CC be used for drug screening procedures. The present sequence represents a  
CC Streptococcus pneumoniae nucleic acid of the invention. Note: The  
CC sequence data for this patent did not appear in the printed specification  
CC but was obtained in electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html.  
XX XX  
XX SQ Sequence 2325 BP; 725 A; 513 C; 461 G; 626 T; 0 U; 0 Other;  
Query Match 73.6%; Score 16.2; DB 13; Length 2325;  
Best Local Similarity 85.7%; Pred. No. 6.7e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 TCATGCTCACACATTAACTGG 21  
||||| ||||| ||||| |||||  
Db 489 TCATGCACACACTGTAACCTGG 509  
RESULT 29  
ADR92735  
ID ADR92735 standard; DNA; 2343 BP.  
XX AC  
XX ADR92735;  
XX XX  
XX 16-DEC-2004 (first entry)  
XX DE  
XX Novel S. pneumoniae DNA sequence, SEQ ID 1370.  
XX XX  
XX Meningitis; bacteraemia; pneumonia; otitis media; ds;  
XX bacterial infection.  
XX OS  
XX Streptococcus pneumoniae.  
XX OS  
XX US6800744-B1.  
XX PN  
XX 05-OCT-2004.  
XX PD  
XX 30-JUN-1998; 98US-00107433.  
XX PF  
XX PR

PR 02-JUL-1997; 97US-0051553P.  
PR 12-MAY-1998; 98US-0085131P.  
XX (GENO-) GENOME THERAPEUTICS CORP.  
XX PA Doucette-Stamm LA, Bush D;  
XX PI  
XX PI  
XX WPI; 2004-697205/68.  
XX DR P-PSDB; ADR95338.  
XX DR  
XX New isolated nucleic acid encoding a Streptococcus pneumoniae  
PT polypeptide, useful for diagnosing, preventing and/or treating  
PT pathological conditions resulting from the bacterial infection.  
XX XX  
XX PS Disclosure; SEQ ID NO 1370; 151pp; English.  
XX XX  
XX The invention relates to an isolated nucleic acid comprising a sequence  
CC encoding a Streptococcus pneumoniae ADR91366polypeptide, or its  
CC fragments, with any of 9 fully defined sequences (appearing as ADR94308,  
CC ADR94489, ADR94800, ADR94837, ADR94969, ADR95253, ADR95642, ADR95682,  
CC ADR96079) or any of the fully defined sequences appearing as ADR91705,  
CC ADR91886, ADR92197, ADR92234, ADR93039, ADR93079, ADR92366, ADR92650 or  
CC ADR93476 or at least 20 or 30 consecutive nucleotides of the nucleotide  
CC sequences, or at least 40, 60 or 300 consecutive nucleotides, which is  
CC hybridisable under high stringency conditions to the nucleotide sequence.  
CC The nucleic acids and proteins are chosen from 5206 disclosed sequences.  
CC Also included are a recombinant expression vector comprising the isolated  
CC nucleic acid cited above operably linked to a transcription regulatory  
CC element, a cell comprising the recombinant expression vector and a probe  
CC comprising at least 20 consecutive nucleotides of the nucleotide  
CC sequences as cited above. The methods and compositions of the present  
CC invention are useful for the diagnosis, prevention and/or treatment of  
CC pathological conditions resulting from bacterial infection by  
CC Streptococcus pneumoniae e.g. pneumonia, bacteraemia, meningitis and  
CC otitis media. The present sequence is one of the 2603 disclosed S.  
CC pneumoniae nucleic acid sequences. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=6800744B1.  
XX XX  
XX SQ Sequence 2343 BP; 736 A; 514 C; 464 G; 629 T; 0 U; 0 Other;  
Query Match 73.6%; Score 16.2; DB 13; Length 2343;  
Best Local Similarity 85.7%; Pred. No. 6.7e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 TCATGCTCACACATTAACTGG 21  
||||| ||||| ||||| |||||  
Db 507 TCATGCACACACTGTAACCTGG 527  
RESULT 30  
AEA56605  
ID AEA56605 standard; DNA; 2343 BP.  
XX AC  
XX AEA56605;  
XX XX  
XX 25-AUG-2005 (first entry)  
XX DE  
XX Streptococcus pneumoniae ORF nucleic acid sequence SEQ ID NO:1370.  
XX XX  
XX bacterial infection; Streptococcus pneumoniae infection; antibacterial;  
XX vaccine; gene; ds.  
XX OS  
XX Streptococcus pneumoniae.  
XX OS  
XX US2005136404-A1.  
XX PN  
XX 23-JUN-2005.  
XX PD  
XX 10-JUL-2003; 2003US-00617320.  
XX PF  
XX 02-JUL-1997; 97US-0051553P.  
XX PR

```

PR 12-MAY-1998; 98US-0085131P.
PR 30-JUN-1998; 98US-00107433.
XX
XX
PA (DOUC/) DOUCETTE-STAMM L A.
PA (BUSH/) BUSH D.
XX
XX Doucette-Stamm LA, Bush D;
PI
XX WPI; 2005-477576/48.
DR P-PSDB; AEA59208.
XX
XX New isolated nucleic acid molecules and encoded polypeptides useful for
PT diagnosing, preventing or treating bacterial infections, particularly
PT Streptococcus pneumoniae infection.
XX
XX Claim 1; SEQ ID NO 1370; 144pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule for detecting,
CC preventing or treating pathological conditions resulting from bacterial
CC infection. The isolated nucleic acid comprises: (a) any of the 2603
CC nucleotide sequences of AEA55236 to AEA57838; (b) a nucleotide sequence
CC encoding a Streptococcus pneumoniae polypeptide comprising any of the
CC 2603 amino acid sequences of AEA57839 to AEA60441; or (c) a nucleotide
CC sequence of at least 8 nucleotides in length, where the sequence is
CC hybridizable to a nucleic acid having any of the nucleotide sequences in
CC (a). Also described: (1) a recombinant expression vector comprising the
CC above nucleic acid operably linked to a transcription regulatory element;
CC (2) a cell comprising the recombinant expression vector; (3) producing an
CC S. pneumoniae polypeptide; (4) a probe comprising a nucleotide sequence
CC consisting of at least 8 nucleotides of any of AEA55236 to AEA57838; (5)
CC treating a subject for S. pneumoniae infection; (6) a recombinant or
CC substantially pure preparation of an S. pneumoniae polypeptide or its
CC fragment, where the polypeptide is selected from AEA57839 to AEA60441;
CC (7) a vaccine composition for preventing or treating an S. pneumoniae
CC infection, comprising an amount of the above nucleic acid or polypeptide;
CC (8) detecting the presence of a Streptococcus nucleic acid in a sample;
CC (9) a computer readable medium having recorded the nucleotide sequences
CC of AEA55236 to AEA57838; (10) a computer based system for identifying
CC fragments of the Streptococcus genome of commercial importance. The
CC composition and methods are useful for diagnosing, preventing or treating
CC bacterial infections, particularly S. pneumoniae infection. The present
CC sequence represents a S. pneumoniae ORF nucleic acid sequence from the
CC present invention. Note - The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from the USPTO web site.
XX
XX Sequence 2343 BP; 736 A; 514 C; 464 G; 629 T; 0 U; 0 Other;
SQ
Query Match 73.6%; Score 16.2; DB 14; Length 2343;
Best Local Similarity 85.7%; Pred. No. 6.7e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TCATGCTCACACATTAAGTGG 21
Db 507 TCATGCACACACTGTAAGTGG 527

RESULT 31
ABN68555
ID AEN68555 standard; DNA; 2352 BP.
XX
XX AEN68555;
AC
XX 01-JUL-2002 (first entry)
DT
XX Streptococcus polynucleotide SEQ ID NO 5023.
DE
XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
XX Streptococcus pyogenes.
OS
XX

```

---

```

PN WO200234771-A2.
XX
XX 02-MAY-2002.
XX
XX 29-OCT-2001; 2001WO-GB004789.
XX
XX 27-OCT-2000; 2000GB-00026333.
PR 24-NOV-2000; 2000GB-00028727.
PR 07-MAR-2001; 2001GB-00005640.
XX
XX (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
XX Telford J, Maignani V, Margarit Y Rosi, Grandi G, Fraser C;
PI Tettelin H;
XX
XX WPI: 2002-352536/38.
DR P-PSDB; ABP27924.
XX
XX New Streptococcus protein for the treatment or prevention of infection or
PT disease caused by Streptococcus bacteria, such as meningitis, and for
PT detecting a compound that binds to the protein.
XX
XX Claim 7; Page 3664; 4525pp; English.
XX
XX The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (1), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins
XX
XX Sequence 2352 BP; 710 A; 514 C; 466 G; 662 T; 0 U; 0 Other;
SQ
Query Match 73.6%; Score 16.2; DB 6; Length 2352;
Best Local Similarity 85.7%; Pred. No. 6.7e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TCATGCTCACACATTAAGTGG 21
Db 510 TCATGCCACACTGTAAGTGG 530

RESULT 32
AAZ97024/c
ID AAZ97024 standard; cDNA; 2374 BP.
XX
XX AAZ97024;
AC
XX 19-APR-2000 (first entry)
DT
XX Human secreted protein gene 6 cDNA clone HCE3G20, SEQ ID NO:16.
DE
XX Human; secreted protein; cancer; tumour; developmental abnormality;
KW foetal deficiency; blood disorder; immune system disorder; inflammation;
KW autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;
KW schizophrenia; arthritis; asthma; psoriasis; sepsis; skin disorder;
KW atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;
KW digestive disorder; endocrine disorder; infection; AIDS; leukaemia;
KW therapy; chromosome 2; ds.
XX
XX Homo sapiens.
OS
XX

```

```

XX WO9966041-A1.
PN
XX
XX 23-DEC-1999.
PD
XX
XX 15-JUN-1999; 99WO-US013418.
PF
XX
XX 16-JUN-1998; 98US-0089507P.
PR
XX 16-JUN-1998; 98US-0089508P.
PR
XX 16-JUN-1998; 98US-0089509P.
PR
XX 16-JUN-1998; 98US-0089510P.
PR
XX 22-JUN-1998; 98US-0090112P.
PR
XX 22-JUN-1998; 98US-0090113P.
PR
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Ruben SM, Ni J, Rosen CA, Wei Y, Young PE, Florence KA;
PI Soppet DR, Brewer LA, Endress GA, Carter KC, Mucenski M, Ebner R;
PI Lafleur DW, Olsen HS, Shi Y, Moore PA, Komatsoulis G;
XX
XX WPI; 2000-106100/09.
DR
XX P-PSDB; AAY86220.
DR
XX
XX New isolated human genes and the secreted polypeptides they encode,
PT useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders.
XX
XX Claim 1; Page 319; 586pp; English.
PS
XX
XX AAZ97019 to AAZ97137 represent 94 isolated human secreted protein genes.
CC
XX AAY86215 to AAY86333 are the secreted proteins encoded by the 94 human
CC genes. This sequence represents a fragment of one of the human secreted
CC proteins. The genes and their corresponding secreted polypeptides are
CC useful for preventing, treating or ameliorating medical conditions, e.g.,
CC by protein or gene therapy. Also pathological conditions can be diagnosed
CC by determining the amount of the new polypeptides in a sample or by
CC determining the presence of mutations in the new genes. Specific uses are
CC described for each of the 94 genes, based on which tissues they are most
CC highly expressed in, and include developing products for the diagnosis or
CC treatment of cancer, tumours, developmental abnormalities and foetal
CC deficiencies, blood disorders, diseases of the immune system, autoimmune
CC diseases, inflammation, allergies, Alzheimer's and cognitive disorders,
CC schizophrenia, arthritis, asthma, psoriasis, sepsis, skin disorders,
CC atherosclerosis, diabetes, cardiovascular disorders, kidney disorders,
CC digestive/endocrine disorders, infections and AIDS. The polypeptides are
CC also useful for identifying their binding partners. The sequences shown
CC in AAY86334 to AAY86585 represent fragments of the secreted proteins
XX
XX Sequence 2374 BP; 672 A; 458 C; 483 G; 743 T; 0 U; 18 Other;
SQ
Query Match 73.5%; Score 16.2; DB 3; Length 2374;
Best Local Similarity 85.7%; Pred. No. 6.7e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTAACTGGA 22
Db |||||
532 CAGCTTCACACATTAACTGGA 512

RESULT 33
ACH66653/c
ID ACH66653 standard; cDNA; 2374 BP.
XX
XX ACH66653;
AC
XX
XX 06-NOV-2003 (first entry)
DT
XX
XX Novel human secreted protein #6 cDNA.
DE
XX
XX Human; ss; vaccine; immune system disorder; haematopoietic cell disorder;
KW cancer; autoimmune disorder; rheumatoid arthritis; glomerulonephritis;
KW HIV infection; anaemia; thrombocytopenia blood coagulation disorder;
KW blood platelet disorder; wound; heart attack; myocardial infarction;
KW

```

```

KW stroke; scarring; asthma; graft-versus host rejection; inflammation;
KW hyperproliferative disorder; lymphoproliferative disorder; arrhythmia;
KW aberrant cellular division; cell proliferative disorder; angiogenesis;
KW cardiovascular disorder; pulmonary heart disease; neovascularisation;
KW hypertrophic scar; keloid; ocular disorder; diabetic retinopathy; gene;
KW uveitis; epithelial cell proliferation; neurological disease; apoptosis;
KW Parkinson's disease; Alzheimer's disease; Huntington's chorea; ALS; AIDS;
KW amyotrophic lateral sclerosis; toxin induced liver disease; septic shock;
KW cachexia; anorexia; lung damage; infection.
XX
XX Homo sapiens.
OS
XX US2003065151-A1.
PN
XX
XX 03-APR-2003.
PD
XX
XX 04-APR-2002; 2002US-00115123.
PF
XX
XX 16-JUN-1998; 98US-0089507P.
PR
XX 16-JUN-1998; 98US-0089508P.
PR
XX 16-JUN-1998; 98US-0089509P.
PR
XX 16-JUN-1998; 98US-0089510P.
PR
XX 22-JUN-1998; 98US-0090112P.
PR
XX 22-JUN-1998; 98US-0090113P.
PR
XX 15-JUN-1999; 99WO-US013418.
PR
XX 14-DEC-1999; 99US-00461325.
PR
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Ruben SM, Ni J, Rosen CA, Wei Y, Young P, Florence K, Soppet DR;
PI Brewer LA, Endress GA, Carter KC, Mucenski M, Ebner R, Lafleur DW;
PI Olsen H, Shi Y, Moore PA, Komatsoulis G;
XX
XX WPI; 2003-531736/50.
DR
XX P-PSDB; ABO53363.
DR
XX
XX Novel antibody that binds specifically to a HCEJQ69 protein, useful for
PT detecting the presence of a protein in a biological sample, and for
PT treating cancers, autoimmune disorders and HIV infection.
XX
XX Disclosure; SEQ ID NO 16; 176pp; English.
XX
XX The invention relates to an isolated antibody or its fragment that
XX specifically binds to a protein. The antibody is useful for detecting a
XX protein in a biological sample, by contacting the biological sample with
XX the antibody or its fragment and detecting the protein in the biological
XX sample. The antibody is useful for purifying, detecting and targeting the
XX human secreted proteins, including both in vitro and in vivo diagnostic
XX and therapeutic methods. The antibody is useful for immunophenotyping of
XX cell lines in biological samples and in antibody-based therapies for
XX treating, inhibiting and preventing diseases, disorders or conditions
XX associated with aberrant expression and/or activity of the above
XX proteins. The antibody is useful for treating deficiencies or disorders
XX of immune system and haematopoietic cells, for increasing differentiation
XX and proliferation of haematopoietic cells, for treating immune
XX deficiencies or disorders e.g. cancers, autoimmune disorders (such as
XX rheumatoid arthritis and glomerulonephritis), HIV infection, anaemia and
XX thrombocytopenia and as a marker for a particular immune system disease
XX or disorder. The antibody is also useful for treating blood coagulation
XX disorders, blood platelet disorders, wounds, heart attacks (infarction),
XX strokes, scarring and asthma. The antibody is also useful for treating or
XX preventing graft-versus host rejection, for modulating inflammatory disorders
XX treating hyperproliferative disorders e.g. lymphoproliferative disorders
XX and cancers, for inhibiting aberrant cellular division and for treating
XX cardiovascular disorders e.g. pulmonary heart disease and arrhythmia,
XX disorders associated with neovascularisation and angiogenesis, for
XX treating hypertrophic scars and keloids, ocular disorders e.g. diabetic
XX retinopathy and uveitis, for wound healing and disorders of epithelial
XX cell proliferation. The antibody is also useful for treating neurological
XX diseases e.g. Parkinson's disease, Alzheimer's disease, Huntington's
XX chorea and amyotrophic lateral sclerosis (ALS), diseases associated with
XX increased apoptosis e.g. AIDS, toxin induced liver disease, septic shock,
XX

```





Db 944 TCCTGGTCACTCAATTAAGTGG 924

RESULT 36  
ADJ96274  
ID ADJ96274 standard; DNA; 3559 BP.  
XX  
XX  
AC ADJ96274;  
XX  
XX 06-MAY-2004 (first entry)  
XX  
XX Human breast cancer-1 DNA #6.  
XX  
XX Breast cancer-1; diagnosis; hyperproliferative disorder; cancer;  
KW antiseize therapy; human; gene; ds.  
XX  
XX Homo sapiens.  
XX  
XX  
FH Key Location/Qualifiers  
FT CDS 142..2184  
FT /\*tag= a  
FT /product= "Human breast cancer-1 protein"  
XX  
XX US2004014051-A1.  
XX  
XX 22-JAN-2004.  
XX  
XX 18-JUL-2002; 2002US-00199676.  
XX  
XX 18-JUL-2002; 2002US-00199676.  
XX  
XX (ISIS-) ISIS PHARM INC.  
XX  
XX Brown-Driver VL, Dobie KW;  
PI  
DR WPI; 2004-121557/12.  
DR P-PSDB; ADJ96473.  
DR GENBANK; NM007298.  
XX  
XX New antisense oligonucleotide compounds, useful for diagnosing,  
PT preventing and/or treating conditions with aberrant activity of breast  
PT cancer-1, such as breast, ovary, prostate and/or peritoneum cancers.  
PT  
XX Example 15; SEQ ID NO 15; 175pp; English.  
XX  
XX The present invention is directed to novel antisense compounds targetted  
CC to breast cancer-1 proteins and their encoding nucleic acids. The  
CC invention is useful for the diagnosis, prevention and/or treatment of  
CC diseases and conditions associated with aberrant expression and activity  
CC of breast cancer-1 such as a hyperproliferative disorder in particular  
CC breast, ovary, prostate and peritoneum cancers. The invention is also  
CC used in antisense therapy. The present sequence is human breast cancer-1  
CC DNA.  
XX  
XX Sequence 3559 BP; 1072 A; 763 C; 827 G; 897 T; 0 U; 0 Other;  
SQ

Query Match 73.6%; Score 16.2; DB 12; Length 3559;  
Best Local Similarity 85.7%; Pred. No. 7.1e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTAAGTGA 22  
|||||  
Db 1914 CATGCCACAGATCAACTGGA 1934

RESULT 37  
ADJ06107  
ID ADX06107 standard; DNA; 3559 BP.  
XX  
XX  
AC ADX06107;  
XX  
XX 21-APR-2005 (first entry)  
DT  
XX

DE Cyclin-dependent kinase modulation biomarker DNA SEQ ID NO 672.  
XX  
KW cytostatic; cyclin-dependent kinase; cdk; biomarker; gene; ds.  
XX  
OS Homo sapiens.  
XX  
XX WO2005012875-A2.  
XX  
XX 10-FEB-2005.  
XX  
XX 29-JUL-2004; 2004WO-US024424.  
XX  
XX 29-JUL-2003; 2003US-0490890P.  
XX  
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
XX  
XX Li M, Rupnow BA, Webster KR, Jackson DG, Wong TW;  
PI  
XX WPI; 2005-163068/17.  
DR P-PSDB; ADX06108.  
XX  
XX Biomarkers useful for predicting or determining the response of a mammal  
PT to a cancer treatment comprising administration of a modulator of cyclin-  
PT dependent kinase activity.  
XX  
XX Claim 5; SEQ ID NO 672; 141pp; English.  
XX  
XX This invention describes a novel method of predicting or determining  
CC whether a mammal will respond or is responding to an anti-cancer agent  
CC that modulates cyclin-dependent kinase (cdk) activity. The method  
CC comprises measuring the level of one or more biomarkers selected from  
CC 2774 biomarkers given in the specification (nucleotide sequence SEQ ID  
CC NO:1246 (Genbank EST W28729), is especially preferred). The method of the  
CC invention is utilized in a kit for determining or predicting whether  
CC patient would be susceptible or resistant to treatment by an agent  
CC modulating cdk activity. The invention also describes a method for  
CC utilizing individualized genetic profiles for treating diseases and  
CC disorders based on patient's response and molecular level, specialized  
CC against the biomarkers and a cell culture model to identify biomarkers.  
CC The cdk modulator is preferably N-5-[(5-(1,1-Dimethylethyl)-2-  
CC oxazolyl)methyl]thiol-2-thiazolyl-4-piperidine carboxamide, 0.5-L-  
CC tartaric acid salt. Note: The sequence data for this patent did not form  
CC part of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences. This  
CC sequence encodes a biomarker used in the method of the invention.  
XX  
XX Sequence 3559 BP; 1072 A; 763 C; 827 G; 897 T; 0 U; 0 Other;  
SQ

Query Match 73.6%; Score 16.2; DB 14; Length 3559;  
Best Local Similarity 85.7%; Pred. No. 7.1e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTAAGTGA 22  
|||||  
Db 1914 CATGCCACAGATCAACTGGA 1934

RESULT 38  
ADW98062  
ID ADW98062 standard; DNA; 3559 BP.  
XX  
XX AC ADW98062;  
XX  
XX 21-APR-2005 (first entry)  
DT  
XX Human breast cancer-1 DNA, seqid:15.  
XX  
XX Antisense therapy; pharmaceutical; gene therapy; diagnostic;  
KW breast tumor; endocrine disease; gynecology and obstetrics; neoplasm;  
KW infection; antimicrobial; inflammation; antineoplastic; tumor;  
KW cytostatic; cancer; breast cancer-1; BRCA1; chromosome 17; gene; ds.  
XX



```
OS Homo sapiens.
XX Key Location/Qualifiers
FH 142..2184
FT /*tag= a
FT /product= "Breast cancer-1 protein"
XX
PN US2005026857-A1.
XX
XX 03-FEB-2005.
XX
XX 12-AUG-2003; 2003US-00639300.
XX
XX 18-JUL-2002; 2002US-00199676.
XX
XX (BROW/) BROWN-DRIVER V L.
PA (DOBI/) DOBIE K W.
XX
XX Brown-Driver VL, Dobie KW;
XX WPI; 2005-141410/15.
DR P-PSDB; ADW98137.
DR REFSEQ; NM_007298.1.
XX
XX New antisense oligonucleotide compound having a specified length of
XX nucleic acid that hybridizes to a human breast cancer-1 sequence, useful
XX for the treatment of breast cancer.
XX
XX Example 15; SEQ ID NO 15; 176pp; English.
XX
XX The present invention relates to a compounds, compositions and methods
XX for modulating the expression of human breast cancer-1. The composition
XX comprises antisense oligonucleotides targeted to nucleic acid encoding
XX breast cancer-1. Breast cancer-1 is also known as breast cancer 1, early
XX onset; BRCA-1; BRCA1; papillary serous carcinoma of peritoneum (PSCP).
XX The invention is useful for the treatment of breast cancer and to prevent
XX delay infection, inflammation or tumor formation. The present sequence is
XX the human breast cancer-1 DNA. The breast cancer-1 gene is located on
XX chromosome 17q2-q21.
XX
XX Sequence 3559 BP; 1072 A; 763 C; 827 G; 897 T; 0 U; 0 Other;
SQ
Query Match 73.6%; Score 16.2; DB 14; Length 3559;
Best Local Similarity 85.7%; Pred. No. 7.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 CATGCTCACACATTAACTGGA 22
Db 1914 CATGCCACAGATCAACTGGA 1934
RESULT 39
ADK67836
ID ADK67836 standard; cDNA; 3682 BP.
XX
XX AC ADK67836;
XX
XX 06-MAY-2004 (first entry)
XX
XX Human BRCC120 gene coding sequence.
XX
XX Human; BRCC120; breast cancer; cytostatic;
XX BRCA1-BRCA2-containing complex; BRCC; gene; ss.
XX
XX Homo sapiens.
XX Key Location/Qualifiers
FH 142..2307
FT /*tag= a
FT /product= "BRCC120"
XX
XX WO2004012755-A1.
XX
```

```
PD 12-FEB-2004.
XX
XX 05-AUG-2003; 2003WO-US024394.
XX
XX 05-AUG-2002; 2002US-0401433P.
XX 24-FEB-2003; 2003US-0449950P.
XX
XX (WIST-) WISTAR INST.
XX
XX Shiekhattar R;
XX
XX WPI; 2004-180395/17.
DR P-PSDB; ADK67825.
XX
XX Modulating the activity of at least one component of a BRCA1-BRCA2-
XX containing complex (BRCC), for treating BRCC-associated cancer, comprises
XX by contacting BRCC with an agent that interacts with a BRCC36 or BRE
XX nucleic acid.
XX
XX Disclosure; SEQ ID NO 18; 144pp; English.
XX
XX The present sequence comprises the coding sequence of the human BRCC120
XX gene. Using stable cell lines expressing epitope tagged BARD1 (BRCA1-
XX associated RING domain), a multiprotein complex containing BRCA2, BRCA1
XX and RAD51 ADK67819-ADK67821, as well as BRCC300, BRCC140, BRCC130,
XX BRCC120, BRCC80, BRE, BRCC36 and BARD1 ADK67822-ADK67829 was identified.
XX This complex, termed the BRCA1-BRCA2-containing complex (BRCC), was found
XX to have an associated ubiquitin E3 ligase activity which was regulated by
XX the negative regulator, BRCC36, as well as DNA repair and transcriptional
XX regulator activities. The present invention relates to methods for
XX regulating at least one component of the BRCC. The method involves
XX contacting BRCC or a cell containing BRCC with an agent that interacts
XX with a nucleic acid sequence encoding BRCC36 or BRE, or a product
XX thereof, thereby altering an activity of BRCC. Methods of identifying
XX such agents and using them to treat cancer are also provided. BRCC36-
XX specific antibodies and methods for diagnosing cancers associated with
XX BRCC are also described. Cancers associated with BRCC include breast,
XX ovarian, prostate, and colon cancer. Polynucleotides encoding the BRCC
XX components can be used for recombinant production of the BRCC components
XX polypeptides.
XX
XX Sequence 3682 BP; 1114 A; 784 C; 855 G; 929 T; 0 U; 0 Other;
SQ
Query Match 73.6%; Score 16.2; DB 12; Length 3682;
Best Local Similarity 85.7%; Pred. No. 7.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 CATGCTCACACATTAACTGGA 22
Db 2037 CATGCCACAGATCAACTGGA 2057
RESULT 40
ADJ96279
ID ADJ96279 standard; DNA; 3682 BP.
XX
XX AC ADJ96279;
XX
XX 06-MAY-2004 (first entry)
XX
XX Human breast cancer-1 DNA #11.
XX
XX Breast cancer-1; diagnosis; hyperproliferative disorder; cancer;
XX antisense therapy; human; gene; ds.
XX
XX Homo sapiens.
XX Key Location/Qualifiers
FH 142..2307
FT /*tag= a
FT /product= "Human breast cancer-1 protein"
XX
XX US2004014051-A1.
XX
```

XX 22-JAN-2004.  
PD 18-JUL-2002; 2002US-00199676.  
XX PF 18-JUL-2002; 2002US-00199676.  
XX PR 18-JUL-2002; 2002US-00199676.  
XX PA (ISIS-) ISIS PHARM INC.  
XX PI Brown-Driver VL, Dobie KW;  
XX WPI; 2004-121557/12.  
DR P-PSDB; ADJ96478.  
DR GENBANK; NM007103.  
XX PT New antisense oligonucleotide compounds, useful for diagnosing,  
PT preventing and/or treating conditions with aberrant activity of breast  
PT cancer-1, such as breast, ovary, prostate and/or peritoneum cancers.  
XX PS Example 15; SEQ ID NO 20; 175pp; English.  
XX CC The present invention is directed to novel antisense compounds targetted  
CC to breast cancer-1 proteins and their encoding nucleic acids. The  
CC invention is useful for the diagnosis, prevention and/or treatment of  
CC diseases and conditions associated with aberrant expression and activity  
CC of breast cancer-1 such as a hyperproliferative disorder in particular  
CC breast, ovary, prostate and peritoneum cancers. The invention is also  
CC used in antisense therapy. The present sequence is human breast cancer-1  
CC DNA.  
XX SQ Sequence 3682 BP; 1114 A; 784 C; 855 G; 929 T; 0 U; 0 Other;  
  
Query Match 73.6%; Score 16.2; DB 12; Length 3682;  
Best Local Similarity 85.7%; Pred. No. 7.1e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 2 CATGCTCACACATTAACTGGA 22  
Db 2037 CATGCCACAGATCAACTGGA 2057  
  
Search completed: November 20, 2005, 17:43:53  
Job time : 360.432 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2005, 16:52:01 ; Search time 3290.22 Seconds  
(without alignments)  
312.841 Million cell updates/sec

Title: US-10-627-757-28

Perfect score: 22  
Sequence: 1 tcatgtcacacattactgga 22

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_hc:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_est7:\*  
9: gb\_gss1:\*  
10: gb\_gss2:\*  
11: gb\_gss3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	22	100.0	382	2	BE087354	QV1-BT0680
C 2	22	100.0	494	1	AW271623	x81603.x
C 3	22	100.0	529	1	AW271514	x815c03.x
C 4	19.4	88.2	2301	10	AY407040	AY407040 Homo sapi
C 5	19.4	88.2	2301	10	AY407041	AY407041 Pan trogl
C 6	18.8	85.5	271	5	BU447141	BU447141 603765767
C 7	18.8	85.5	577	2	BF907433	MR4-UT005
C 8	18.8	85.5	702	5	BU142547	603136317
C 9	18.8	85.5	819	5	BU229136	603947883
C 10	18.8	85.5	863	5	BU340622	603519652
C 11	18.8	85.5	879	2	BE658217	GM700005A
C 12	18.4	83.6	583	2	BF616038	HVSMSC000
C 13	18.4	81.8	637	9	AZ405422	IM0174C13
C 14	18	81.8	711	10	CE368446	tigr-gss-
C 15	18	81.8	1394	8	CV945942	PV007A1 z
C 16	17.8	80.9	71	10	CG506716	OST56377
C 17	17.8	80.9	119	2	BE475856	946049B09
C 18	17.8	80.9	242	2	BE518656	946073A06
C 19	17.8	80.9	332	3	BQ357514	QV0-HT080
C 20	17.8	80.9	408	2	BF809824	CMI-C1013
C 21	17.8	80.9	505	9	AQ182879	HS 3113_B
C 22	17.8	80.9	582	9	CC518811	CH240_365

23	17.8	80.9	595	7	CO608095	CO608095 Dg8-6c18
C 24	17.8	80.9	601	10	CG177305	CG177305 PUICK88TD
C 25	17.8	80.9	605	9	BZ223274	BZ223274 CH230-260
C 26	17.8	80.9	607	9	BZ676221	BZ676221 PUBCW73TD
C 27	17.8	80.9	615	10	BX226160	BX226160 Danilo rer
C 28	17.8	80.9	628	10	BX159437	BX159437 Danilo rer
C 29	17.8	80.9	641	10	CG359130	CG359130 OG48G22TC
C 30	17.8	80.9	666	2	BE970982	BE970982 601651082
C 31	17.8	80.9	705	8	CX720060	CX720060 1326860 N
C 32	17.8	80.9	709	9	CC340540	CC340540 OG7AD73TV
C 33	17.8	80.9	726	10	AG342302	AG342302 Mus muscu
C 34	17.8	80.9	735	8	CX720061	CX720061 1326861 N
C 35	17.8	80.9	736	10	AG443132	AG443132 Mus muscu
C 36	17.8	80.9	792	10	CG338883	CG338883 OGMWH74TH
C 37	17.8	80.9	795	10	CG372456	CG372456 OGVVDV70TV
C 38	17.8	80.9	805	11	CR898868	CR898868 Sus scrof
C 39	17.8	80.9	823	10	DU071829	DU071829 63619 Tom
C 40	17.8	80.9	828	10	BX231022	BX231022 Danilo rer
C 41	17.8	80.9	842	8	DN056873	DN056873 JGI-CABAI
C 42	17.8	80.9	850	10	CZ985095	CZ985095 202792 To
C 43	17.8	80.9	982	2	BG481475	BG481475 602528857
C 44	17.8	80.9	1044	9	BZ784018	BZ784018 PUGBH28TB
C 45	17.8	80.9	1167	10	CL083101	CL083101 ISBI-1G12

ALIGNMENTS

RESULT 1  
BE087354/c  
LOCUS  
DEFINITION  
QV1-BT0680-240400-166-f02 BT0680 Homo sapiens cDNA, mRNA linear EST 12-JUN-2000  
ACCESSION  
BE087354  
VERSION  
BE087354.1 GI:8477749  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
REFERENCE  
1 (bases 1 to 382)  
AUTHORS  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.  
TITLE  
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
10737800  
COMMENT  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=QV1-BT0680-240400-166-f02&t3=2000-04-24&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 9  
High quality sequence stop: 382.  
Location/Qualifiers  
1 . 382  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="BT0680"

/note="Organ: breast; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

## ORIGIN

Query Match 100.0%; Score 22; DB 2; Length 382;  
Best Local Similarity 100.0%; Pred. No. 8.3;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCATGCTCACATTAACCTGGA 22

DB 235 TCATGCTCACATTAACCTGGA 214

## RESULT 2

AW271623 494 bp mRNA linear EST 03-JAN-2000  
LOCUS xs16e03.x1 NCI\_CGAP\_Kid11 Homo sapiens cDNA clone IMAGE:2769820 3',  
DEFINITION mRNA sequence.

ACCESSION AW271623.1 GI:6658653

VERSION AW271623

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

REFERENCE 1 (bases 1 to 494)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.

DNA Sequencing by: Greg Lennon, Ph.D.

Cloned through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco

High quality sequence stop: 464.

Location/Qualifiers

1..494

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:2769820"

/lab\_host="DH10B"

/clone\_lib="NCI\_CGAP\_Kid11"

/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with

a modified polylinker; Site 1: Not I; Site 2: Eco RI;

Plasmid DNA from the normalized library NCI\_CGAP Kid3 was

prepared, and ss circles were made in vitro. Following HAP

purification, this DNA was used as tracer in a subtractive

hybridization reaction. The driver was PCR-amplified cDNAs

from a pool of 5,000 clones made from the same library

(cloneIDs 1322376-1323911, 1456007-1456775, and

1500552-1502855). Subtraction by Bento Soares and M.

Fatima Bonaldo."

ORIGIN

Query Match 100.0%; Score 22; DB 1; Length 494;  
Best Local Similarity 100.0%; Pred. No. 8.6;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCATGCTCACATTAACCTGGA 22

Db 160 TCATGCTCACATTAACCTGGA 181

## RESULT 3

AW271514

LOCUS

DEFINITION

xs15c03.x1 NCI\_CGAP\_Kid11 Homo sapiens cDNA clone IMAGE:2769700 3',

mRNA sequence.

ACCESSION AW271514

VERSION AW271514.1 GI:6658544

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

REFERENCE 1 (bases 1 to 529)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.

DNA Sequencing by: Greg Lennon, Ph.D.

Cloned through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco

High quality sequence stop: 458.

Location/Qualifiers

1..529

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:2769700"

/lab\_host="DH10B"

/clone\_lib="NCI\_CGAP\_Kid11"

/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with

a modified polylinker; Site 1: Not I; Site 2: Eco RI;

Plasmid DNA from the normalized library NCI\_CGAP Kid3 was

prepared, and ss circles were made in vitro. Following HAP

purification, this DNA was used as tracer in a subtractive

hybridization reaction. The driver was PCR-amplified cDNAs

from a pool of 5,000 clones made from the same library

(cloneIDs 1322376-1323911, 1456007-1456775, and

1500552-1502855). Subtraction by Bento Soares and M.

Fatima Bonaldo."

ORIGIN

Query Match 100.0%; Score 22; DB 1; Length 529;

Best Local Similarity 100.0%; Pred. No. 8.6;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCATGCTCACATTAACCTGGA 22

DB 160 TCATGCTCACATTAACCTGGA 181

## RESULT 4

AY407040

LOCUS

DEFINITION

Homo sapiens SIM1 gene, VIRTUAL TRANSCRIPT, partial sequence,

genomic survey sequence.

ACCESSION AY407040

VERSION AY407040.1 GI:39763011

KEYWORDS GSS.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens





## AUTHORS

Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,  
Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.  
A Comprehensive Collection of Chicken cDNAs  
Curr. Biol. 12 (22), 1965-1969 (2002)

## PUBMED

12445392

## COMMENT

Contact: Simon Hubbard  
Department of Biomolecular Sciences  
University of Manchester Institute of Science and Technology  
(UMIST)  
PO Box 88, Manchester, M60 1QD, UK

Tel: 01612008930

Fax: 01612360409

Email: Simon.Hubbard@umist.ac.uk.

## FEATURES

## source

1. .819  
Location/Qualifiers  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/strain="White Leghorn, Hisex"  
/db\_xref="taxon:9031"  
/clone="CHEST903e22"  
/dev\_stage="22"  
/lab\_host="DH10B"  
/clone\_lib="CSEQCHN23"

/note="Organ: head; Vector: pBluescript II KS(+); Site\_1:  
EcoRI; Site\_2: NotI; This normalized library was  
constructed from 1 million independent clones. cDNA  
synthesis was initiated using an oligo(dT) primer, using  
methylated C in the first strand synthesis reaction.  
Following this first strand reaction, double-stranded cDNA  
was bluntended, ligated to NotI adapters, digested with  
EcoRI, size-selected, and cloned into the NotI and EcoRI  
compatible sites of a custom modified MCS of the  
pBluescript (KS+) vector. The library was normalized in 2  
rounds using conditions adapted from Soares et al., PNAS  
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6  
(1996) 791, except that a significantly longer  
reannealing hybridization was used."

## ORIGIN

Query Match 85.5%; Score 18.8; DB 5; Length 819;

Best Local Similarity 90.9%; Pred. No. 3.3e+02;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCATGCTCACACATTAACTGGA 22

||||| ||||| ||||| |||||

Db 697 TCATGCTGACAGTAACCTGGA 718

## RESULT 10

## BU340622

## LOCUS

BU340622 863 bp mRNA linear EST 28-NOV-2002  
603519652F1 CSEQCHN67 Gallus gallus cDNA clone Chest460h13 5', mRNA  
sequence.

## ACCESSION

## BU340622

## VERSION

## EST.

## KEYWORDS

## SOURCE

## ORGANISM

## Gallus gallus

## Gallus gallus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.

## REFERENCE

## 1 (bases 1 to 863)

## AUTHORS

## Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.

## TITLE

## A Comprehensive Collection of Chicken cDNAs

## JOURNAL

## Curr. Biol. 12 (22), 1965-1969 (2002)

## PUBMED

## 12445392

## COMMENT

Contact: Simon Hubbard  
Department of Biomolecular Sciences  
University of Manchester Institute of Science and Technology  
(UMIST)  
PO Box 88, Manchester, M60 1QD, UK  
Tel: 01612008930

## Fax: 01612360409

## Email: Simon.Hubbard@umist.ac.uk.

## Location/Qualifiers

## 1. .863

## /organism="Gallus gallus"

## /mol\_type="mRNA"

## /strain="White Leghorn, Hisex"

## /db\_xref="taxon:9031"

## /clone="CHEST460h13"

## /dev\_stage="16 day embryo"

## /lab\_host="DH10B"

## /clone\_lib="CSEQCHN67"

/note="Organ: brain; Vector: pBluescript II KS(+); Site\_1:  
EcoRI; Site\_2: NotI; This normalized library was  
constructed from 1 million independent clones. cDNA  
synthesis was initiated using an oligo(dT) primer, using  
methylated C in the first strand synthesis reaction.  
Following this first strand reaction, double-stranded cDNA  
was bluntended, ligated to NotI adapters, digested with  
EcoRI, size-selected, and cloned into the NotI and EcoRI  
compatible sites of a custom modified MCS of the  
pBluescript (KS+) vector. The library was normalized in 2  
rounds using conditions adapted from Soares et al., PNAS  
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6  
(1996) 791, except that a significantly longer  
reannealing hybridization was used."

## FEATURES

## source

1. .863  
Location/Qualifiers

/organism="Gallus gallus"

/mol\_type="mRNA"

/strain="White Leghorn, Hisex"

/db\_xref="taxon:9031"

/clone="CHEST460h13"

/dev\_stage="16 day embryo"

/lab\_host="DH10B"

/clone\_lib="CSEQCHN67"

/note="Organ: brain; Vector: pBluescript II KS(+); Site\_1:  
EcoRI; Site\_2: NotI; This normalized library was  
constructed from 1 million independent clones. cDNA  
synthesis was initiated using an oligo(dT) primer, using  
methylated C in the first strand synthesis reaction.  
Following this first strand reaction, double-stranded cDNA  
was bluntended, ligated to NotI adapters, digested with  
EcoRI, size-selected, and cloned into the NotI and EcoRI  
compatible sites of a custom modified MCS of the  
pBluescript (KS+) vector. The library was normalized in 2  
rounds using conditions adapted from Soares et al., PNAS  
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6  
(1996) 791, except that a significantly longer  
reannealing hybridization was used."

## ORIGIN

Query Match 85.5%; Score 18.8; DB 5; Length 863;

Best Local Similarity 90.9%; Pred. No. 3.4e+02;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCATGCTCACACATTAACTGGA 22

||||| ||||| ||||| |||||

Db 784 TCATGCTGACAGTAACCTGGA 805

## RESULT 11

## BE658217

## LOCUS

BE658217 879 bp mRNA linear EST 24-MAY-2001  
GM700005A10E5 Gm-r1070 Glycine max cDNA clone Gm-r1070-1737 3',  
mRNA sequence.

## ACCESSION

## BE658217

## VERSION

## EST.

## KEYWORDS

## SOURCE

## Glycine max (soybean)

## ORGANISM

## Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
Glycine.

## REFERENCE

## 1 (bases 1 to 879)

## AUTHORS

Vodkin,L., Keim,P., Shoemaker,R., Retzel,E., Khanna,A., Coryell,V.,  
Erpelding,J., Raph,C., Shoop,E., Pardinas,J., Liu,L. and Lewin,H.

## TITLE

## A Functional Genomics Program for Soybean (NSF 9872565)

## JOURNAL

## Unpublished (1999)

## COMMENT

## Other\_ESTs: AF941033 corresponding to Gm-cl010-1595 (5')

## Contact: Vodkin, L.O., PI, A Functional Genomics Program for

## Soybean (NSF 9872565)

## Lewin, H. A., Director, Keck Center for Comparative and Functional

## Genomics

## University of Illinois

## Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA

## Tel: (217) 244-6147

## Fax: (217) 333-4582

## Email: l-vodkin@uiuc.edu

## This clone is available through: Genome Systems, Inc. 4633 World

## Parkway Circle St. Louis, Missouri 63134. For further information

## call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)

## 427-3324 or contact: clones@genomesystems.com or info@genome

## systems.com web site: www.genomesystems.com

## Seq primer: 5'-TTTTTTTTTTTTTTT(A/C/G)-3'.

## Location/Qualifiers

source

```

1. .879
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="Gm-r1070-1737"
/clone_lib="Gm-r1070"
/note="The library Gm-r1070 is a sequence-driven, reracked
set of 9,216 clones selected from cDNA libraries from
various tissues and stages of development of soybean that
represent 2,639 sequences from immature cotyledons, 1,770
from immature seed coats, 3,938 from flowers, and 869
from young pods. The 5' ESTs of the source clones from
the different libraries was used to select singletons, or
a representative of each contig, which were reracked to
form library Gm-r1070. The cDNA clones of the reracked
Gm-r1070 library were then sequenced at the 3' end. The
contig analysis to select unique genes was performed by
the laboratory of Ernest Retzel, Center for Computational
Genomics and Bioinformatics, University of Minnesota,
http://www.cbc.umn.edu/ResearchProjects/soybean/index.html
. Reracking was performed by Genome Systems, St. Louis,
http://www.genomesystems.com, and 3' sequencing by the
Kecck Center for Comparative and Functional Genomics,
University of Illinois,
http://www.life.uiuc.edu/biotech/keck.html. Note: The
corresponding 5' EST from each clone in the Gm-r1070
library is listed in the 'OTHER EST' field. The detailed
information on the source library for each clone can also
be obtained by referring to the Genome Systems clone ID of
the original cDNA library that is also listed under
'OTHER EST'."

ORIGIN
Query Match 85.5%; Score 18.8; DB 2; Length 879;
Best Local Similarity 90.9%; Pred. No. 3.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCATGCTCACATTAACCTGGA 22
|||||
DB 700 TCATGCTCACATAAACTTGA 721

RESULT 12
BF616038 583 bp mRNA linear EST 22-OCT-2001
LOCUS
DEFINITION
HVSMEC0001D01f Hordeum vulgare seedling shoot EST library
HVCDNA0003 (Etiolated and unstressed) Hordeum vulgare subsp.
vulgare cDNA clone HVSMEC0001D01f, mRNA sequence.

ACCESSION
BF616038
VERSION
BF616038.2 GI:13106474
KEYWORDS
Hordeum vulgare subsp. vulgare
SOURCE
Hordeum vulgare subsp. vulgare
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.
REFERENCE
1 (bases 1 to 583)
Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D.,
Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W.,
Fenton,R.D., Oates,R. and Main,D.
Development of a genetically and physically anchored EST resource
for barley genomics: Morex unstressed seedling shoot cDNA library
Unpublished (2001)
On Dec 18, 2000 this sequence version replaced gi:11879772.
Contact: Wing RA
Clemson University
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hg bases = 105
Seq primer: AATTAACCTCTACTAAAGGG

```

FEATURES  
source

High quality sequence stop: 581.

```

Location/Qualifiers
1..583
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Morex"
/sub_species="vulgare"
/db_xref="taxon:112509"
/clone="HVSMEC0001D01f"
/tissue_type="Seedling shoot"
/lab_host="TJCI21"
/clone_lib="Hordeum vulgare seedling shoot EST library
HVCDNA0003 (Etiolated and unstressed)"
/note="Vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI.
Seeds were surface sterilized then germinated under axenic
conditions in the dark at room temperature on filter paper
with water, nystatin and cefotaxime in covered
crystallization dishes. Five-day old seedling shoots were
then harvested, total RNA was prepared, poly(A) RNA was
purified, one primary unamplified cDNA library was made,
and 1 million pfu were in vivo excised to give pBluescript
SK(-) cDNA phagemids. These steps were performed in the TJ
Close laboratory at the University of California,
Riverside (Choi, Close, Fenton). Phagemids were plated and
picked at the Clemson University Genomics Institute (CUGI)
(Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA
preparations, DNA sequencing and sequence analysis were
performed at CUGI (Wing, Yu, Frisch, Henry, Simmons,
Oates, Rambo, Main). The sequence has been trimmed to
remove vector sequence and contains a minimum of 100 bases
of phred value 20 or above. For more details on library
preparation and sequence analysis see
http://www.genome.clemson.edu/projects/barley. To order
this clone see http://www.genome.clemson.edu/orders/Also
see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
Genetically and physically anchored EST resources for
barley genomics. Barley Genetics Newsletter 31:29-30.
(http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

```

ORIGIN

Query Match 83.6%; Score 18.4; DB 2; Length 583;  
Best Local Similarity 95.0%; Pred. No. 5.1e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCATGCTCACATTAACCTG 20

DB 158 TCATGCTCACATTAACCTG 177

RESULT 13  
AZ405422/c

LOCUS

DEFINITION

1M0174C13F Mouse 10kb plasmid UUGC1M library Mus musculus genomic

clone UUGC1M0174C13 F, genomic survey sequence.

ACCESSION

AZ405422

VERSION

AZ405422.1 GI:10529435

KEYWORDS

GSS.

SOURCE

Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridea; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 637)

Dunn,D., Aoyagi,A., Barber,M., Bescorn,T., Duval,B., Hamil,C.,

Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,

Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von

Niederhausern,A. and Wright,D., Weiss,R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah



Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0174 row: C column: 13  
 Seq primer: CGTTGTAACACGACGCCAGT  
 Class: plasmid ends  
 High quality sequence stop: 637.

# FEATURES

Location/Qualifiers  
 1..637  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC1M0174C13"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, Tl-resistant, P-"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

# ORIGIN

Query Match 81.8%; Score 18; DB 9; Length 637;  
 Best Local Similarity 100.0%; Pred. No. 8e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCATGCTCACATTAAAC 18  
 ||||||||||||||||  
 Db 212 TCATGCTCACATTAAAC 195

RESULT 14  
 CE368446

LOCUS  
 DEFINITION

CE368446 711 bp DNA linear GSS 27-SEP-2003  
 tigr-gss-dog-17000362034306 Dog Library Canis familiaris genomic,  
 genomic survey sequence.

ACCESSION  
 VERSION

KEYWORDS  
 SOURCE

ORGANISM

Canis familiaris (dog)

Canis familiaris  
 Canis familiaris  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;  
 Canis.

REFERENCE  
 AUTHORS

1 (bases 1 to 711)  
 Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,  
 Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and  
 Venter,J.C.

TITLE  
 JOURNAL

PUBMED  
 COMMENT

The dog genome: survey sequencing and comparative analysis  
 Science 301 (5641), 1898-1903 (2003)  
 14512627

Contact: Kirkness EF  
 The Institute for Genomic Research  
 Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,

# FEATURES

source  
 Location/Qualifiers  
 1..711  
 /organism="Canis familiaris"  
 /mol\_type="genomic DNA"  
 /strain="Standard Poodle"  
 /db\_xref="taxon:9615"  
 /clone\_lib="Dog Library"  
 /note="Site 1: BstXI; Libraries were prepared from peripheral blood"

# ORIGIN

Query Match 81.8%; Score 18; DB 10; Length 711;  
 Best Local Similarity 100.0%; Pred. No. 8.1e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CATGCTCACATTAACT 19  
 ||||||||||||||||  
 Db 337 CATGCTCACATTAACT 354

RESULT 15  
 CV945942

LOCUS  
 DEFINITION

CV945942 1394 bp mRNA linear EST 25-JAN-2005  
 PV007AI zoospores, purified Phytophthora infestans cDNA, mRNA  
 sequence.

ACCESSION  
 VERSION

KEYWORDS

SOURCE

ORGANISM

Phytophthora infestans (potato late blight agent)

Phytophthora infestans  
 Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;  
 Phytophthora.

REFERENCE  
 AUTHORS

1 (bases 1 to 1394)  
 Randall,T., Dwyer,R.A., Huitema,E., Beyer,K., Cvitanich,C.,  
 Kelkar,H., Fong,A.M., Gates,K., Roberts,S., Yatzkan,E., Gaffney,T.,  
 Law,M., Testa,A., Torto-Alalibo,A., Zhang,M., Zheng,L., Mueller,E.,  
 Windass,J., Binder,A., Birch,P.R.J., Giel,U., Govers,F., Gow,N.A.,  
 Mauch,F., van West,P., Waugh,M.E., Yu,J., Boller,T., Kamoun,S.,  
 Lam,S.T. and Judelson, H.S.  
 Large-scale gene discovery in the oomycete Phytophthora infestans  
 reveals likely components of phytopathogenicity shared with true  
 fungi

Phytophthora infestans (potato late blight agent)

Phytophthora infestans  
 Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;  
 Phytophthora.

JOURNAL  
 PUBMED

COMMENT

Mol. Plant-Microbe Interact. 18 (3), 229-243 (2005)  
 15782637  
 Contact: Judelson HS  
 Department of Plant Pathology  
 University of California  
 Webber Hall, Riverside, CA 92521, USA  
 Tel: 909 787 4199  
 Fax: 909 787 4294  
 Email: howard.judelson@ucr.edu.

# FEATURES

source  
 Location/Qualifiers  
 1..1394  
 /organism="Phytophthora infestans"  
 /mol\_type="mRNA"  
 /strain="88069"  
 /db\_xref="taxon:4787"  
 /sex="AI"  
 /clone\_lib="zoospores, purified"  
 /note="Vector: pSPORT1"

# ORIGIN

Query Match 81.8%; Score 18; DB 8; Length 1394;  
 Best Local Similarity 100.0%; Pred. No. 8.7e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CATGCTCACATTAACT 19  
 ||||||||||||||||

```

Db      720 CATGCTCACACATTAAC 737

RESULT 16
LOCUS   CG506716/c
DEFINITION 71 bp mRNA linear GSS 01-OCT-2003
OST56377 Mus musculus 129Sv/Ev Mus musculus cDNA clone OST56377,
mRNA sequence.
ACCESSION CG506716
VERSION   CG506716.1 GI:37286104
KEYWORDS GSS.
SOURCE   Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
1 (bases 1 to 71)
REFERENCE
AUTHORS  Zambrowicz,B.P., Abuin,A., Ramirez-Solis,R., Richter,L.J.,
Piggott,J., BeltrandelRio,H., Buxton,E.C., Edwards,J., Finch,R.A.,
Fridde,C.J., Gupta,A., Hansen,G., Hu,Y., Huang,W., Jaing,C.,
Key,B.W. Jr., Kipp,P., Kohlhaufl,B., Ma,Z.-Q., Markesich,D.,
Payne,R., Potter,D.G., Qian,N., Shaw,J., Schrick,J., Shi,Z.-Z.,
Sparks,M.J., Van Slichtenhorst,I., Vogel,P., Walke,W., Xu,N.,
Zhu,Q., Person,C. and Sands,A.T.
TITLE    Wnk1 kinase deficiency lowers blood pressure in mice: a gene-trap
screen to identify potential targets for therapeutic intervention
JOURNAL  Proc. Natl. Acad. Sci. U.S.A. 100 (24), 14109-14114 (2003)
PUBMED   14610273
COMMENT  Contact: Zambrowicz BP
OmniBank
Lexicon Genetics Incorporated
4000 Research Forest Drive, The Woodlands, TX 77381, USA
Email: materials@lexgen.com
Gene trap sequence tag generated by 3' RACE from mouse ES cells as
described in Zambrowicz et al (Nature. 1998 Apr 9;392(6676):608-11)
Class: Gene Trap.
FEATURES             Location/Qualifiers
     source
     1..71
     /organism="Mus musculus"
     /mol_type="mRNA"
     /strain="129Sv/Ev"
     /db_xref="taxon:10090"
     /clone="OST56377"
     /cell_type="embryonic stem cell"
     /clone_lib="Mus musculus 129Sv/Ev"

ORIGIN
Query Match      80.9%; Score 17.8; DB 10; Length 71;
Best Local Similarity 90.5%; Pred. No. 8.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TCATGCTCACACATTAACGG 21
        |||||
        24 TCATGCTCACATCTTAACGG 4

Db

RESULT 17
LOCUS   BE475856
DEFINITION 119 bp mRNA linear EST 28-JUL-2000
946049B09.x2 946 - tassell primordium prepared by Schmidt lab Zea
mays cDNA, mRNA sequence.
ACCESSION BE475856
VERSION   BE475856.1 GI:9566347
KEYWORDS EST.
SOURCE   Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 119)
REFERENCE
AUTHORS  Walbot,V.
TITLE    Maize ESTs from various cDNA libraries sequenced at Stanford
University

Db      720 CATGCTCACACATTAAC 737

RESULT 18
LOCUS   BE518656/c
DEFINITION 242 bp mRNA linear EST 08-AUG-2000
946073A06.y1 946 - tassell primordium prepared by Schmidt lab Zea
mays cDNA, mRNA sequence.
ACCESSION BE518656
VERSION   BE518656.1 GI:9742493
KEYWORDS EST.
SOURCE   Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 242)
REFERENCE
AUTHORS  Walbot,V.
TITLE    Maize ESTs from various cDNA libraries sequenced at Stanford
University
JOURNAL  Unpublished (1999)
COMMENT  Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 946073 row: A column: 06.
Location/Qualifiers
     source
     1..242
     /organism="Zea mays"
     /mol_type="mRNA"
     /cultivar="OH43"
     /db_xref="taxon:4577"
     /tissue_type="tassels"
     /dev_stage="just after the transition from vegetative to
inflorescence development"
     /lab_host="XL0LR"
     /clone_lib="946 - tassell primordium prepared by Schmidt
lab"
     /note="Organ: tassels; Vector: HybridZAP; Site 1: EcoRI;
Site 2: XhoI; George Chuck dissected immature tassels
between 1mm and 3mm. Sharon Stanfield prepared the cDNA
library in HybridZAP. Sample insert size range was 350 bp
to 3 Kb with a 1 Kb average."

ORIGIN
Query Match      80.9%; Score 17.8; DB 2; Length 119;
Best Local Similarity 90.5%; Pred. No. 8.6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 CATGCTCACATTAACCTGGA 22
        |||||
        29 CATGAGCACATTAACCTGGA 49

Db

RESULT 19
LOCUS   BE518656
DEFINITION 242 bp mRNA linear EST 08-AUG-2000
946073A06.y1 946 - tassell primordium prepared by Schmidt lab Zea
mays cDNA, mRNA sequence.
ACCESSION BE518656
VERSION   BE518656.1 GI:9742493
KEYWORDS EST.
SOURCE   Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 242)
REFERENCE
AUTHORS  Walbot,V.
TITLE    Maize ESTs from various cDNA libraries sequenced at Stanford
University
JOURNAL  Unpublished (1999)
COMMENT  Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 946073 row: A column: 06.
Location/Qualifiers
     source
     1..242
     /organism="Zea mays"
     /mol_type="mRNA"
     /cultivar="OH43"
     /db_xref="taxon:4577"
     /tissue_type="tassels"
     /dev_stage="just after the transition from vegetative to
inflorescence development"
     /lab_host="XL0LR"

```



Unpublished (2003)  
Other GSSs: CH240\_365H23.TARBAC13P2  
Contact: Rob Holt  
Sequencing

The British Columbia Cancer Agency Genome Science Centre  
600 W. 10th Ave, Vancouver, British Columbia, Canada V5Z 4E6  
Tel: 604-877-6085  
Fax: 604-877-6276  
Email: rholt@bccgsc.ca  
Clones are derived from the bovine BAC library CHOR1-240  
(<http://www.chori.org/bacpac/bovine240.htm>). For BAC library  
availability, please contact Pieter de Jong (pdejong@mail.cho.org).  
Clones may be purchased from BACPAC Resources.

was undertaken as part of the International Bovine BAC Mapping Consortium (IBBMC) by CSIRO Livestock Industries, Australia and the

British Columbia Genome Sciences Centre, Canada.  
Plate: 365 row: H column: 23  
Seq primer: T7

```

Class: BAC ends.
      Location/Qualifiers
      1..582
      /organism="Bos taurus"
      /mol_type="genomic DNA"
      /strain="breed: Hereford"
      /db_xref="taxon:9913"
      /clone="CH240_365H23"
      /sex="Male"
      /cell_type="Blood"
      /clone_lib="CHORI-240"
      /note="Vector: pTARBAC1.3; Site 1: MboI; site 2: MboI;

```

```

      100% 80.9% ; Score 17.8 ; DB 9 ; Length 582 ;
      1 Similarity 90.5% ; Pred. No. 1e+03 ;
      19 ; Conservative 0 ; Mismatches 0 ; Indels 0 ; Gaps 0 ;
      2 CATGCTCACACATTAACCTGGA 22
          |||||

```

56 CATGCTCCACATTAACTGAA 36

CO6080895 595 bp mRNA linear EST 21-JUL-2004  
 D88-6c18 D88-testis Canis familiaris cDNA 3', mRNA sequence.  
 CO6080895  
 CO6080895.1 GI:50453659

EST.  
Canis familiaris (dog)  
Canis familiaris  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;  
Canis.  
1 (bases 1 to 595)

Schriener, T., Hermanns, J., Weindel, M., Schuette, D., Kranz, H.,  
 Henrich, J. and Loebbert, R.  
 Dog arrayTAG cDNA clone collection  
 Unpublished (2004)  
 Contact: Thomas Schlueter  
 LION bioscience AG  
 Walldorferstrasse 98, D-69123 Heidelberg, Germany  
 Tel.: +49 6221 4038 150

Fax: +49 6221 4038 290  
Email: Thomas.Schluter@ionbioscience.com.  
Location/Qualifiers  
1. .595  
/organism="Canis familiaris"  
/mol\_type="mrna"  
/strain="Beagle"  
/db\_xref="taxon:9615"

```

/tissue_type="testis"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="DG8-testis"
/notes="Organ: testis; Vector: Dog pBluescript LION"

ORIGIN
Query Match      80.9%; Score 17.8; DB 7; Length 595;
Best Local Similarity 90.5%; Pred. No. 1e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCATGCTCACACATTAACTGG 21
|||||
Db 557 TCATGCTCACACGTTAAACAGG 577

RESULT 24
CG177305/C
LOCUS PUICK88TD_ZM_0.6_1.0_KB_Zea_mays_genomic_clone_ZMMBTA0551P07,
DEFINITION genomic survey sequence.
ACCESSION CG177305
VERSION CG177305.1 GI:34068363
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 601)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
Bennetzen,J.
TITLE Maize Genomics Consortium
JOURNAL Unpublished (2003)
COMMENT Other_GSSs: PUICK88TB
Contact: Cathy Whitelaw
TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
FEATURES
Location/Qualifiers
source
1..601
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBTA0551P07"
/clone_lib="ZM_0.6_1.0_KB"
/notes="Vector: pCR4-T0FO; Site:1: EcoRI; 0.6-1.0 kb high
Cot selected genomic DNA library"

ORIGIN
Query Match      80.9%; Score 17.8; DB 10; Length 601;
Best Local Similarity 90.5%; Pred. No. 1e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCATGCTCACACATTAACTGG 21
|||||
Db 459 TCATGCTCAACATTAACTGG 439

RESULT 25
BZ223274
LOCUS CH230-260C20.TJ CHORI-230 Segment 2 Rattus norvegicus genomic clone
DEFINITION CH230-260C20, genomic survey sequence.
ACCESSION BZ223274
VERSION BZ223274.1 GI:23881632
KEYWORDS GSS.

Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 605)
AUTHORS Zhao,S., Shetty,J., Shatsman,S., Tsagay,G., Geer,K.,
Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D.,
Riggs,F., de Jong,P. and Fraser,C.M.
TITLE Rat BAC End Sequences from Library CHORI-230 MboI segment
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: CH230-260C20.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat230.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or ering_information.htm). BAC end
page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Plate: 260 row: C column: 20
Seq primer: SP6
Class: BAC ends.
FEATURES
Location/Qualifiers
source
1..605
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SaNHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-260C20"
/sex="female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 2"
/notes="Vector: pTARBAC1.3; Site 1: MboI; Site 2: MboI;
CHORI-230 Rat (BN/SaNHsd/MCW) BAC library produced by
Pieter de Jong"

ORIGIN
Query Match      80.9%; Score 17.8; DB 9; Length 605;
Best Local Similarity 90.5%; Pred. No. 1e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCATGCTCACACATTAACTGG 21
|||||
Db 444 TCATGCACACACATTCACTGG 464

RESULT 26
BZ676221
LOCUS PUBCW73TD_ZM_0.6_1.0_KB_Zea_mays_genomic_clone_ZMMBTA026N01,
DEFINITION genomic survey sequence.
ACCESSION BZ676221
VERSION BZ676221.1 GI:28227656
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 607)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
Bennetzen,J.
TITLE Maize Genomics Consortium
JOURNAL Unpublished (2003)
COMMENT Contact: Cathy Whitelaw
TIGR

```





```

/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone_lib="ZMMBMA0387N02"
/clone_lib="ZM.0.7.1.5 KB"
/note="Vector: pBCSK-1; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Query Match      80.9%; Score 17.8; DB 9; Length 709;
Best Local Similarity 90.5%; Pred. No. 1e+03; 2; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 0;

QY 2 CATGCTCACACATTAAGTGA 22
    ||||| ||||| ||||| |||||
DB 73 CATGACACACATTAAGTGA 93

RESULT 33
AG342302/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus molossinus (Japanese wild mouse)

REFERENCE
AUTHORS
Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
Shiroishi,T.
TITLE
Contribution of Asian mouse subspecies Mus musculus molossinus to
genomic constitution of strain C57BL/6J, as defined by BAC-end
sequence-SNP analysis
JOURNAL
PUBMED
Genome Res. 14 (12), 2439-2447 (2004)
15574823
2 (bases 1 to 726)
Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
Direct Submission
Submitted (17-NOV-2003) Mashira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
(E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the mouse BAC library MSMg01. For BAC
library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
Tsukuba Institute, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : TJ
LIBRARY
Vector : pBACe3.6
R.Site 1 : ECORI
R.Site 2 : EcoRI.
Location/Qualifiers
1. .726
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/adb_species="molossinus"
/adb_xref="taxon:57486"
/clone_lib="MSMg01-135P16.TJ"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"

FEATURES
source
AG342302
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus molossinus (Japanese wild mouse)

REFERENCE
AUTHORS
Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
Shiroishi,T.
TITLE
Contribution of Asian mouse subspecies Mus musculus molossinus to
genomic constitution of strain C57BL/6J, as defined by BAC-end
sequence-SNP analysis
JOURNAL
PUBMED
Genome Res. 14 (12), 2439-2447 (2004)
15574823
2 (bases 1 to 726)
Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
Direct Submission
Submitted (17-NOV-2003) Mashira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
(E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the mouse BAC library MSMg01. For BAC
library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
Tsukuba Institute, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : TJ
LIBRARY
Vector : pBACe3.6
R.Site 1 : ECORI
R.Site 2 : EcoRI.
Location/Qualifiers
1. .726
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/adb_species="molossinus"
/adb_xref="taxon:57486"
/clone_lib="MSMg01-135P16.TJ"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"

ORIGIN

```

```

/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone_lib="ZMMBMA0387N02"
/clone_lib="ZM.0.7.1.5 KB"
/note="Vector: pBCSK-1; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Query Match      80.9%; Score 17.8; DB 10; Length 726;
Best Local Similarity 90.5%; Pred. No. 1e+03; 2; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 0;

QY 2 CATGCTCACACATTAAGTGA 22
    ||||| ||||| ||||| |||||
DB 625 CATGCTCAACATTAAGTGA 605

RESULT 34
CX720061/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Oncorhynchus mykiss (rainbow trout)

REFERENCE
AUTHORS
Rexroad,C.E., Govoroun,M., Le Gac,F., Guiguen,Y. and Yao,J.
TITLE
O9RT testis, NCCCWA/WVU EST Project, Phase II, in collaboration
with INRA
JOURNAL
Unpublished (2005)
COMMENT
Contact: Rexroad CE
USDA, ARS, National Center for Cool and Cold Water Aquaculture
11876 Leetown Road, Kearneysville, WV 25430, USA
Tel: 304 724 8340 x2129
Fax: 304 725 0351
Email: crexroad@nccwa.ars.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and
trimmed with the aid of the trim_alt option. Vector identified with
cross_match v0.990329.
Plate: 1 row: K column: 8
Seq primer: ATTTAGGTGACACTATAG.
Location/Qualifiers
1. .735
/organism="Oncorhynchus mykiss"
/mol_type="mRNA"
/db_xref="taxon:8022"
/adb_host="DH10B"
/clone_lib="NCCCWA O9RT"
/note="Vector: pT7T3D-PacI; INRA testis library made by M.
Govoroun, F. Le Gac, and Y. Guiguen."

FEATURES
source
CX720061
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Oncorhynchus mykiss (rainbow trout)

REFERENCE
AUTHORS
Rexroad,C.E., Govoroun,M., Le Gac,F., Guiguen,Y. and Yao,J.
TITLE
O9RT testis, NCCCWA/WVU EST Project, Phase II, in collaboration
with INRA
JOURNAL
Unpublished (2005)
COMMENT
Contact: Rexroad CE
USDA, ARS, National Center for Cool and Cold Water Aquaculture
11876 Leetown Road, Kearneysville, WV 25430, USA
Tel: 304 724 8340 x2129
Fax: 304 725 0351
Email: crexroad@nccwa.ars.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and
trimmed with the aid of the trim_alt option. Vector identified with
cross_match v0.990329.
Plate: 1 row: K column: 8
Seq primer: ATTTAGGTGACACTATAG.
Location/Qualifiers
1. .735
/organism="Oncorhynchus mykiss"
/mol_type="mRNA"
/db_xref="taxon:8022"
/adb_host="DH10B"
/clone_lib="NCCCWA O9RT"
/note="Vector: pT7T3D-PacI; INRA testis library made by M.
Govoroun, F. Le Gac, and Y. Guiguen."

ORIGIN
Query Match      80.9%; Score 17.8; DB 8; Length 735;
Best Local Similarity 90.5%; Pred. No. 1e+03; 2; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 0;

QY 2 CATGCTCACACATTAAGTGA 22
    ||||| ||||| ||||| |||||
DB 413 CATACACACATTAAGTGA 393

RESULT 35
AG443132
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus molossinus (Japanese wild mouse)

REFERENCE
AUTHORS
Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
Shiroishi,T.
TITLE
Contribution of Asian mouse subspecies Mus musculus molossinus to
genomic constitution of strain C57BL/6J, as defined by BAC-end
sequence-SNP analysis
JOURNAL
PUBMED
Genome Res. 14 (12), 2439-2447 (2004)
15574823
2 (bases 1 to 726)
Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
Direct Submission
Submitted (17-NOV-2003) Mashira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
(E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the mouse BAC library MSMg01. For BAC
library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
Tsukuba Institute, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : TJ
LIBRARY
Vector : pBACe3.6
R.Site 1 : ECORI
R.Site 2 : EcoRI.
Location/Qualifiers
1. .726
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/adb_species="molossinus"
/adb_xref="taxon:57486"
/clone_lib="MSMg01-135P16.TJ"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"

ORIGIN

```



Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and Shiroishi,T.  
 Contribution of Asian mouse subspecies *Mus musculus molossinus* to genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SNP analysis  
 Genome Res. 14 (12), 2439-2447 (2004)  
 15574823  
 2 (bases 1 to 736)  
 Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.  
 Direct Submission  
 Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC) and 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170  
 Clones are derived from the mouse BAC library MSMg01. For BAC library availability, please contact Kuniya Abe (abe@rtc.riken.jp).  
 Tsukuba Institute, Bio Resource Center.  
 The Institute of Physical and Chemical Research (RIKEN) 3-1-1 Koyadai, Tsukuba, 305-0074 Japan  
 phone: 81-298-36-9189, fax: 81-298-36-9199  
 e-mail: abe@rtc.riken.jp  
 PRIMERS  
 Sequencing : TJ

LIBRARY  
 Vector : pBACe3.6  
 R.Site 1 : EcoRI  
 R.Site 2 : EcoRI.

#### FEATURES

source  
 1..736  
 Location/Qualifiers

/organism="Mus musculus molossinus"  
 /mol\_type="genomic DNA"  
 /sub\_species="molossinus"  
 /db\_xref="taxon:57486"  
 /clone="MSMg01-325C19.TJ"  
 /sex="male"  
 /tissue\_type="mixture of kidney and spleen"  
 /clone\_lib="MSMg01 Mouse Male BAC Library"

#### ORIGIN

Query Match 80.9%; Score 17.8; DB 10; Length 736;  
 Best Local Similarity 90.5%; Pred. No. 1e+03; 2; Indels 0; Gaps 0;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCATGCTCACACATTAACTGG 21

|||||  
 133 TCATGCTCACACATTAGCGG 153

#### RESULT 36

CG338883  
 LOCUS  
 DEFINITION  
 OGWH74TH ZM\_0.7\_1.5\_KB Zea mays genomic clone ZMMEMa0616N04,  
 genomic survey sequence.

ACCESSION  
 VERSION  
 CG338883.1 GI:34256149  
 GSS.

#### SOURCE

Ze mays  
 Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 792)

#### REFERENCE

AUTHORS  
 Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,  
 Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,  
 Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.  
 Consortium for Maize Genomics  
 Unpublished (2002)

Other GSSs: OGWH74TV

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-838-5843

Fax: 301-838-0208  
 Email: whitelaw@tigr.org  
 Seq primer: TF  
 Class: methylation filtered.

#### FEATURES

source  
 1..792  
 Location/Qualifiers  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"  
 /clone="ZMMEMa0616N04"  
 /clone\_lib="ZM\_0.7\_1.5\_KB"  
 /note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb  
 methylation filtered genomic DNA library"

#### ORIGIN

Query Match 80.9%; Score 17.8; DB 10; Length 792;  
 Best Local Similarity 90.5%; Pred. No. 1e+03; 2; Indels 0; Gaps 0;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTAACCTGGA 22

|||||  
 206 CATGAGCACACATTAACCTGGA 226

#### RESULT 37

CG372456  
 LOCUS  
 DEFINITION  
 OGVDV70TV ZM\_0.7\_1.5\_KB Zea mays genomic clone ZMMEMa0518L20,  
 genomic survey sequence.

ACCESSION  
 VERSION  
 CG372456.1 GI:34289723  
 GSS.

#### SOURCE

Ze mays  
 Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 795)

#### AUTHORS

Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,  
 Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,  
 Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.

Consortium for Maize Genomics

Unpublished (2002)

Other GSSs: OGVDV70TH

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TF

Class: methylation filtered.

#### FEATURES

source  
 1..795  
 Location/Qualifiers  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"  
 /clone="ZMMEMa0518L20"  
 /clone\_lib="ZM\_0.7\_1.5\_KB"  
 /note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb  
 methylation filtered genomic DNA library"

#### ORIGIN

Query Match 80.9%; Score 17.8; DB 10; Length 795;  
 Best Local Similarity 90.5%; Pred. No. 1e+03; 2; Indels 0; Gaps 0;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTAACCTGGA 22

|||||  
 403 CATGAGCACACATTAACCTGGA 423

```

RESULT 38
CR898868      805 bp      DNA      linear      GSS 23-NOV-2004
LOCUS
DEFINITION   Sus scrofa BES, genomic survey sequence.
ACCESSION   CR898868
VERSION     CR898868.1  GI:56223365
KEYWORDS    GSS; Bac-end sequence BES; Genome Survey Sequence.
SOURCE      Sus scrofa (pig)
ORGANISM    Sus scrofa
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;
             Sus.
REFERENCE   1 (bases 1 to 805)
AUTHORS    Rogel-Gaillard,C., Bourgeaux,N., Billault,A., Vaiman,M. and
            Chardon,P.
TITLE      Construction of a swine BAC library: application to the
            characterization and mapping of porcine type C endoviral elements
JOURNAL    Cytogenet. Cell Genet. 85 (3-4), 205-211 (1999)
PUBMED    10449899
REFERENCE   2 (bases 1 to 805)
AUTHORS    Chardon,P., Iannuccelli,N., Roig,A., Dossat,C., Denars,J.,
            Rogel-Gaillard,C., Roy,A., Schibler,L. and Milan,D.
TITLE      A physical map of the swine genome
JOURNAL    Unpublished
REFERENCE   3 (bases 1 to 805)
AUTHORS    Genoscope.
TITLE      Direct Submission
JOURNAL    Submitted (18-NOV-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
FEATURES    Location/Qualifiers
             1..805
             /organism="Sus scrofa"
             /mol_type="genomic DNA"
             /strain="Large White"
             /db_xref="taxon:9823"
             /clone="b10254H04"
             /sex="male"
             /cell_type="fibroblast"
             /clone_lib="SBAB"
             /note="Genoscope sequence ID : IH0AAA25CG12FM1"
ORIGIN
Query Match      80.9%; Score 17.8; DB 11; Length 805;
Best Local Similarity 90.5%; Pred. No. 1e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCATGCTCACATTAACTGG 21
    |||||
Db 699 TTATACTCACATTAACTGG 719

RESULT 39
DU071829
LOCUS
DEFINITION   63619 Tomato HindIII BAC Library lycopersicon esculentum genomic
            clone LE HBa0048F10 5, genomic survey sequence.
ACCESSION   DU071829
VERSION     DU071829.1  GI:72507970
KEYWORDS    GSS.
SOURCE      Lycopersicon esculentum (Solanum lycopersicum)
ORGANISM    Lycopersicon esculentum
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
             asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1 (bases 1 to 823)
AUTHORS    Mueller,J.A., Buels,R.M., Wang,Y., Tanksley,S.D., Giovannoni,J.J.,
            Van Eck,J. and Stack,S.
TITLE      BAC end sequencing from three Solanum lycopersicon libraries
JOURNAL    Unpublished (2005)
COMMENT    Other GSSs: 63618
            ContaCt: Lukas Mueller

```

```

Tanksley Lab, Dept. of Plant Breeding
Cornell University
251 Emerson Hall, Ithaca, NY 14853, USA
Tel: 607-255-6557
Fax: 607-255-6683
Email: sgn-feedback@sgn.cornell.edu
Insert Length: 30972 Std Error: 0.00
Plate: 48 row: F column: 10
Seq primer: 17
Class: BAC ends
High quality sequence start: 44
High quality sequence stop: 823.
Location/Qualifiers
1..823
/organism="Lycopersicon esculentum"
/mol_type="genomic DNA"
/cultivar="Heinz 1706"
/db_xref="taxon:4081"
/clone="LE HBa0048F10"
/lab_host="E. coli"
/clone_lib="Tomato HindIII BAC Library"
/note="Vector: pBelOBAC11; Site_1: HindIII"
ORIGIN
Query Match      80.9%; Score 17.8; DB 10; Length 823;
Best Local Similarity 90.5%; Pred. No. 1e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CATGCTCACATTAACTGGA 22
    |||||
Db 267 CATGCTCACACTTAACTGGA 287

RESULT 40
BX231022/c
LOCUS
DEFINITION   Danio rerio genomic clone DKEY-254F7, genomic survey sequence.
ACCESSION   BX231022
VERSION     BX231022.1  GI:28065172
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
ORGANISM    Danio rerio
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
             Cypriniformes; Cyprinidae; Danio.
REFERENCE   1 (bases 1 to 828)
AUTHORS    Humphray,S.J., Huckle,E. and Durham,J.L.
TITLE      Direct Submission
JOURNAL    Submitted (13-MAR-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
            This sequence was generated from the SP6 end of BAC 254F7. 254F7 is
            part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES    Location/Qualifiers
             1..828
             /organism="Danio rerio"
             /mol_type="genomic DNA"
             /db_xref="taxon:7955"
             /clone="DKEY-254F7"
             /tissue_type="Testis"
             /note="Vector pIndigoBAC-536"
ORIGIN
Query Match      80.9%; Score 17.8; DB 10; Length 828;
Best Local Similarity 90.5%; Pred. No. 1e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCATGCTCACATTAACTGG 21
    |||||
Db 424 TCATGCACACATTAAATGG 404

```

Search completed: November 20, 2005, 21:55:38  
Job time : 3293.22 secs

---

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2005, 17:14:50 ; Search time 100.222 Seconds  
(without alignments)  
390.196 Million cell updates/sec

Title: US-10-627-757-28  
Perfect score: 22  
Sequence: 1 tcattgctcacacattactgga 22

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA:  
1: /cgn2\_6/ptodata/1/ina/1\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/H\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq:\*  
7: /cgn2\_6/ptodata/1/ina/PP\_COMB.seq:\*  
8: /cgn2\_6/ptodata/1/ina/RE\_COMB.seq:\*  
9: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	16.8	76.4	152070	3	US-09-949-016-15402, A
C 2	16.2	73.6	601	3	US-09-949-016-169253, Sequence 169253,
C 3	16.2	73.6	601	3	US-09-949-016-169254, Sequence 169254,
C 4	16.2	73.6	2325	3	US-09-583-110-2042, Sequence 2042, Ap
C 5	16.2	73.6	2343	3	US-09-107-433-1370, Sequence 1370, Ap
C 6	16.2	73.6	2374	3	US-09-461-325-16, Sequence 16, Appl
C 7	16.2	73.6	2374	3	US-10-012-542-16, Sequence 16, Appl
C 8	16.2	73.6	2374	3	US-10-115-123-16, Sequence 16, Appl
C 9	16.2	73.6	4401	3	US-09-614-034-192, Sequence 192, App
C 10	16.2	73.6	4907	3	US-09-949-016-2701, Sequence 2701, Ap
C 11	16.2	73.6	4910	3	US-09-023-655-1125, Sequence 1125, Ap
C 12	16.2	73.6	4910	3	US-09-814-915A-75, Sequence 75, Appl
C 13	16.2	73.6	4910	3	US-09-949-016-129, Sequence 129, App
C 14	16.2	73.6	5656	2	US-08-825-886-1, Sequence 1, Appli
C 15	16.2	73.6	5656	2	US-08-825-886-1, Sequence 1, Appli
C 16	16.2	73.6	5656	3	US-08-989-890-1, Sequence 1, Appli
C 17	16.2	73.6	5689	2	US-08-425-061-3, Sequence 3, Appli
C 18	16.2	73.6	5689	2	US-08-825-886-3, Sequence 3, Appli
C 19	16.2	73.6	5689	3	US-08-989-890-3, Sequence 3, Appli
C 20	16.2	73.6	5707	2	US-08-425-061-11, Sequence 11, Appl
C 21	16.2	73.6	5707	2	US-08-825-886-11, Sequence 11, Appl
C 22	16.2	73.6	5707	3	US-08-989-890-11, Sequence 11, Appl
C 23	16.2	73.6	5709	2	US-08-425-061-2, Sequence 2, Appli
C 24	16.2	73.6	5709	2	US-08-425-061-7, Sequence 7, Appli

25	16.2	73.6	5709	2	US-08-425-061-8	Sequence 8, Appli
26	16.2	73.6	5709	2	US-08-425-061-9	Sequence 9, Appli
27	16.2	73.6	5709	2	US-08-825-886-2	Sequence 2, Appli
28	16.2	73.6	5709	2	US-08-825-886-7	Sequence 7, Appli
29	16.2	73.6	5709	2	US-08-825-886-8	Sequence 8, Appli
30	16.2	73.6	5709	2	US-08-825-886-9	Sequence 9, Appli
31	16.2	73.6	5709	3	US-08-989-890-2	Sequence 2, Appli
32	16.2	73.6	5709	3	US-08-989-890-7	Sequence 7, Appli
33	16.2	73.6	5709	3	US-08-989-890-8	Sequence 8, Appli
34	16.2	73.6	5709	3	US-08-989-890-9	Sequence 9, Appli
35	16.2	73.6	5710	2	US-08-425-061-6	Sequence 6, Appli
36	16.2	73.6	5710	2	US-08-825-886-6	Sequence 6, Appli
37	16.2	73.6	5710	3	US-08-989-890-6	Sequence 6, Appli
38	16.2	73.6	5711	2	US-08-425-061-4	Sequence 4, Appli
39	16.2	73.6	5711	2	US-08-425-061-10	Sequence 10, Appli
40	16.2	73.6	5711	2	US-08-598-591-1	Sequence 1, Appli
41	16.2	73.6	5711	2	US-08-798-691-1	Sequence 1, Appli
42	16.2	73.6	5711	2	US-08-798-691-3	Sequence 3, Appli
43	16.2	73.6	5711	2	US-08-798-691-5	Sequence 5, Appli
44	16.2	73.6	5711	2	US-08-825-886-4	Sequence 4, Appli
45	16.2	73.6	5711	2	US-08-825-886-10	Sequence 10, Appli

ALIGNMENTS

RESULT 1  
US-09-949-016-15402/c  
; Sequence 15402, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15402  
; LENGTH: 152070  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-15402

Query Match 76.4%; Score 16.8; DB 3; Length 152070;  
Best Local Similarity 90.0%; Pred. No. 1.5e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 ATGCTCACACATTAACTGGA 22  
|||||  
Db 131985 ATGTCACAGATTAACTGGA 131966

RESULT 2  
US-09-949-016-169253  
; Sequence 169253, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20

```
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 169253
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-169253

Query Match          73.6%; Score 16.2; DB 3; Length 601;
Best Local Similarity 85.7%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 CATGCTCACACATTAAGTGA 22
        ||||| ||||| |||||
Db      54 CATGCTCGCACAGAAACTGA 74

RESULT 3
US-09-949-016-169254
; Sequence 169254, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 169254
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-169254

Query Match          73.6%; Score 16.2; DB 3; Length 601;
Best Local Similarity 85.7%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 CATGCTCACACATTAAGTGA 22
        ||||| ||||| |||||
Db      317 CATGCTCGCACAGAAACTGA 337

RESULT 4
US-09-583-110-2042
; Sequence 2042, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics
; FILE REFERENCE: PATH00-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
```

```
; SEQ ID NO 2042
; LENGTH: 2325
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-2042

Query Match          73.6%; Score 16.2; DB 3; Length 2325;
Best Local Similarity 85.7%; Pred. No. 1.8e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 TCATGCTCACACATTAAGTGG 21
        ||||| ||||| |||||
Db      489 TCATGCACACACTGTAAGTGG 509

RESULT 5
US-09-107-433-1370
; Sequence 1370, Application US/09107433
; Patent No. 6800744
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; TITLE OF INVENTION: SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
; TITLE OF INVENTION: THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 1370:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2343 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...2343
; SEQUENCE DESCRIPTION: SEQ ID NO: 1370:
US-09-107-433-1370

Query Match          73.6%; Score 16.2; DB 3; Length 2343;
Best Local Similarity 85.7%; Pred. No. 1.8e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

QY 1 TCATGCTCACACATTAAGTGG 21  
||||| ||||| ||||| |||||  
Db 507 TCATGACACACTGTAAGTGG 527

RESULT 6  
US-09-461-325-16/c  
; Sequence 16, Application US/09461325A  
; Patent No. 6475753  
; GENERAL INFORMATION:  
; APPLICANT: Ruben et al.  
; TITLE OF INVENTION: 94 Human Secreted Proteins  
; FILE REFERENCE: P2029P1  
; CURRENT APPLICATION NUMBER: US/09/461,325A  
; CURRENT FILING DATE: 1999-12-14  
; EARLIER APPLICATION NUMBER: PCT/US99/13418  
; EARLIER FILING DATE: 1999-06-15  
; EARLIER APPLICATION NUMBER: 60/089,507  
; EARLIER FILING DATE: 1998-06-16  
; EARLIER APPLICATION NUMBER: 60/089,508  
; EARLIER FILING DATE: 1998-06-16  
; EARLIER APPLICATION NUMBER: 60/089,509  
; EARLIER FILING DATE: 1998-06-16  
; EARLIER APPLICATION NUMBER: 60/089,510  
; EARLIER FILING DATE: 1998-06-16  
; EARLIER APPLICATION NUMBER: 60/090,112  
; EARLIER FILING DATE: 1998-06-22  
; EARLIER APPLICATION NUMBER: 60/090,113  
; EARLIER FILING DATE: 1998-06-22  
; NUMBER OF SEQ ID NOS: 532  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 16  
; LENGTH: 2374  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (556)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: SITE  
; LOCATION: (2344)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: SITE  
; LOCATION: (2346)  
; OTHER INFORMATION: n equals a,t,g, or c  
; US-09-461-325-16

Query Match 73.6%; Score 16.2; DB 3; Length 2374;  
Best Local Similarity 85.7%; Pred. No. 1.8e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTAAGTGG 22  
||| ||||| ||||| |||||  
Db 532 CAGCTTCACACATTAAGTGG 512

RESULT 7  
US-10-012-542-16/c  
; Sequence 16, Application US/10012542  
; Patent No. 6627741  
; GENERAL INFORMATION:  
; APPLICANT: Ruben et al.  
; TITLE OF INVENTION: 94 Human Secreted Proteins  
; FILE REFERENCE: P2029P1  
; CURRENT APPLICATION NUMBER: US/10/012,542  
; CURRENT FILING DATE: 2001-12-12  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/461,325  
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-12-14  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,507  
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,508

; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,509  
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,510  
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/090,112  
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/090,113  
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22  
; NUMBER OF SEQ ID NOS: 532  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 16  
; LENGTH: 2374  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (556)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: SITE  
; LOCATION: (2344)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: SITE  
; LOCATION: (2346)  
; OTHER INFORMATION: n equals a,t,g, or c  
; US-10-012-542-16

Query Match 73.6%; Score 16.2; DB 3; Length 2374;  
Best Local Similarity 85.7%; Pred. No. 1.8e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTAAGTGG 22  
||| ||||| ||||| |||||  
Db 532 CAGCTTCACACATTAAGTGG 512

RESULT 8  
US-10-115-123-16/c  
; Sequence 16, Application US/10115123  
; Patent No. 6774216  
; GENERAL INFORMATION:  
; APPLICANT: Ruben et al.  
; TITLE OF INVENTION: 94 Human Secreted Proteins  
; FILE REFERENCE: P2029G30AP1D2  
; CURRENT APPLICATION NUMBER: US/10/115,123  
; CURRENT FILING DATE: 2002-04-04  
; PRIOR APPLICATION NUMBER: PCT/US99/13418  
; PRIOR FILING DATE: 1999-06-15  
; PRIOR APPLICATION NUMBER: 60/089,507  
; PRIOR FILING DATE: 1998-06-16  
; PRIOR APPLICATION NUMBER: 60/089,508  
; PRIOR FILING DATE: 1998-06-16  
; PRIOR APPLICATION NUMBER: 60/089,509  
; PRIOR FILING DATE: 1998-06-16  
; PRIOR APPLICATION NUMBER: 60/089,510  
; PRIOR FILING DATE: 1998-06-16  
; PRIOR APPLICATION NUMBER: 60/090,112  
; PRIOR FILING DATE: 1998-06-22  
; PRIOR APPLICATION NUMBER: 60/090,113  
; PRIOR FILING DATE: 1998-06-22  
; NUMBER OF SEQ ID NOS: 532  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 16  
; LENGTH: 2374  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (556)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: SITE  
; LOCATION: (2344)  
; OTHER INFORMATION: n equals a,t,g, or c  
; OTHER INFORMATION: n equals a,t,g, or c





```
; APPLICANT: Horwitz, Kathryn
; APPLICANT: Richer, Jennifer
; TITLE OF INVENTION: Progesterone Receptor-Regulated Gene Expression and Methods Relat
; TITLE OF INVENTION: Theteto
; FILE REFERENCE: 2848-39
; CURRENT APPLICATION NUMBER: US/09/814,915A
; CURRENT FILING DATE: 2002-03-21
; PRIOR APPLICATION NUMBER: 60/214,870
; PRIOR FILING DATE: 2000-06-28
; NUMBER OF SEQ ID NOS: 108
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 75
; LENGTH: 4910
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-814-915A-75

Query Match 73.6%; Score 16.2; DB 3; Length 4910;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTAACCTGGA 22
||||| ||||||| |||||
Db 4241 CATGCTGACACATTTCTGGA 4221

RESULT 13
US-09-949-016-129/c
; Sequence 129, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 129
; LENGTH: 4910
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-129

Query Match 73.6%; Score 16.2; DB 3; Length 4910;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTAACCTGGA 22
||||| ||||||| |||||
Db 4241 CATGCTGACACATTTCTGGA 4221

RESULT 14
US-08-425-061-1
; Sequence 1, Application US/08425061
; Patent No. 5622829
; GENERAL INFORMATION:
; APPLICANT: KING, Mary-Claire
; APPLICANT: FRIEDMAN, Lori
; APPLICANT: OSTERMEYER, Beth
; APPLICANT: ROWELL, Sarah
; APPLICANT: LYNCH, Eric
; APPLICANT: SZABO, Csilla
; APPLICANT: LEE, Ming
; TITLE OF INVENTION: GENETIC MARKERS FOR BREAST AND OVARIAN
```

```
; TITLE OF INVENTION: CANCER
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/425,061
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-59563-3/DJB/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5656 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
US-08-425-061-1

Query Match 73.6%; Score 16.2; DB 2; Length 5656;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTAACCTGGA 22
||||| ||||||| |||||
Db 5387 CATGCCACAGATCAACTGGA 5407

RESULT 15
US-08-825-886-1
; Sequence 1, Application US/08825886
; Patent No. 5821328
; GENERAL INFORMATION:
; APPLICANT: KING, Mary-Claire
; APPLICANT: FRIEDMAN, Lori
; APPLICANT: OSTERMEYER, Beth
; APPLICANT: ROWELL, Sarah
; APPLICANT: LYNCH, Eric
; APPLICANT: SZABO, Csilla
; APPLICANT: LEE, Ming
; TITLE OF INVENTION: GENETIC MARKERS FOR BREAST AND OVARIAN
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/825,886
```

;  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/425,061  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OSMAN, Richard A  
; REGISTRATION NUMBER: 36,627  
; REFERENCE/DOCKET NUMBER: A-59563-3/DJB/RAO  
; TELEPHONE: (415) 494-8700  
; TELEFAX: (415) 494-8771  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5656 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
US-08-825-886-1

Query Match 73.6%; Score 16.2; DB 2; Length 5656;  
Best Local Similarity 85.7%; Pred. No. 2e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTAACTGGA 22  
||||| ||||| ||||| ||||| |||||  
DB 5387 CATGCCACAGATCAACTGGA 5407

## RESULT 16

US-08-989-890-1  
; Sequence 1, Application US/08989890  
; Patent No. 6512091

## GENERAL INFORMATION:

; APPLICANT: KING, Mary-Claire  
; FRIEDMAN, Lori  
; OSTERMEYER, Beth  
; ROWELL, Sarah  
; LYNCH, Eric  
; SZABO, Csilla  
; LEE, Ming

TITLE OF INVENTION: GENETIC MARKERS FOR BREAST AND OVARIAN

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 9411

Qy 2 CATGCTCACACATTAAGTGA 22  
||||| ||||| ||||| ||||| |||||  
Db 5419 CATGCCACAGATCAACTGGA 5439

RESULT 18  
US-08-825-886-3  
; Sequence 3, Application US/08825886  
; Patent No. 5821328  
; GENERAL INFORMATION:  
; APPLICANT: KING, Mary-Claire  
; APPLICANT: FRIEDMAN, Lori  
; APPLICANT: OSTERMEYER, Beth  
; APPLICANT: ROWELL, Sarah  
; APPLICANT: LYNCH, Eric  
; APPLICANT: SZABO, Csilla  
; APPLICANT: LEE, Ming  
; TITLE OF INVENTION: GENETIC MARKERS FOR BREAST AND OVARIAN  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/825,886  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/425,061  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OSMAN, Richard A.  
; REGISTRATION NUMBER: 36,627  
; REFERENCE/DOCKET NUMBER: A-59563-3/DJB/RAO  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 494-8700  
; TELEFAX: (415) 494-8771  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5689 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cdna  
US-08-825-886-3

Query Match 73.6%; Score 16.2; DB 2; Length 5689;  
Best Local Similarity 85.7%; Pred. No. 2e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 CATGCTCACACATTAAGTGA 22  
||||| ||||| ||||| ||||| |||||  
Db 5419 CATGCCACAGATCAACTGGA 5439

RESULT 19  
US-08-989-890-3  
; Sequence 3, Application US/08989890  
; Patent No. 6512091  
; GENERAL INFORMATION:  
; APPLICANT: KING, Mary-Claire  
; APPLICANT: FRIEDMAN, Lori  
; APPLICANT: OSTERMEYER, Beth

; ROWELL, Sarah  
; LYNCH, Eric  
; SZABO, Csilla  
; LEE, Ming  
; TITLE OF INVENTION: GENETIC MARKERS FOR BREAST AND OVARIAN  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/989,890  
; FILING DATE: 12-Dec-1997  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/825,886  
; FILING DATE: <Unknown>  
; APPLICATION NUMBER: 08/425,061  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OSMAN, Richard A.  
; REGISTRATION NUMBER: 36,627  
; REFERENCE/DOCKET NUMBER: A-59563-3/DJB/RAO  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 494-8700  
; TELEFAX: (415) 494-8771  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5689 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cdna  
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-08-989-890-3  
  
Query Match 73.6%; Score 16.2; DB 3; Length 5689;  
Best Local Similarity 85.7%; Pred. No. 2e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
Qy 2 CATGCTCACACATTAAGTGA 22  
||||| ||||| ||||| ||||| |||||  
Db 5419 CATGCCACAGATCAACTGGA 5439  
  
RESULT 20  
US-08-425-061-11  
; Sequence 11, Application US/08425061  
; Patent No. 5622829  
; GENERAL INFORMATION:  
; APPLICANT: KING, Mary-Claire  
; APPLICANT: FRIEDMAN, Lori  
; APPLICANT: OSTERMEYER, Beth  
; APPLICANT: ROWELL, Sarah  
; APPLICANT: LYNCH, Eric  
; APPLICANT: SZABO, Csilla  
; APPLICANT: LEE, Ming  
; TITLE OF INVENTION: GENETIC MARKERS FOR BREAST AND OVARIAN  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT  
; STREET: 4 Embarcadero Center, Suite 3400



RESULT 24  
US-08-425-061-7  
; Sequence 7, Application US/08425061  
; Patent No. 5622829  
; GENERAL INFORMATION:  
; APPLICANT: KING, Mary-Claire  
; APPLICANT: FRIEDMAN, Lori  
; APPLICANT: OSTERMEYER, Beth  
; APPLICANT: ROWELL, Sarah  
; APPLICANT: LYNCH, Eric  
; APPLICANT: SZABO, Csilla  
; APPLICANT: LEE, Ming  
; TITLE OF INVENTION: GENETIC MARKERS FOR BREAST AND OVARIAN  
; TITLE OF INVENTION: CANCER  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:

```

? ADDRESS: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
? STREET: 4 Embarcadero Center, Suite 3400
? CITY: San Francisco
? STATE: California
? COUNTRY: USA
? ZIP: 94111-4187
?
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.30
?
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/425,061
? FILING DATE:
? CLASSIFICATION: 435
?

```

ATTORNEY/AGENCY INFORMATION:  
NAME: OSMAN, Richard A  
REGISTRATION NUMBER: 36, 627  
REFERENCE/DOCKET NUMBER: A-59563-3/DJB/RAO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 494-8700  
TELEFAX: (415) 494-8771  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5709 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-08-425-061-7

```
US-08-425-061-7      CDNA
MOLECULE TYPE:
Query Match          73.6%;   Score 16.2;   DB 2;   Length 5709;
Best Local Similarity 85.7%;   Pred. No. 2+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0

QY    2 CATGCTCACACATTAACTGGA 22
      |||||
Db     5439 CATSCCCACATCAACTGGA 5459
```

? APPLICANT: FRIEDMAN, LOUI  
 ? APPLICANT: OSTERMEYER, Beth  
 ? APPLICANT: ROWELL, Sarah  
 ? APPLICANT: LYNCH, Eric  
 ? APPLICANT: SZABO, Csilla  
 ? APPLICANT: LEE, Ming  
 ? TITLE OF INVENTION: GENETIC MARKERS FOR BREAST AND OVARIAN  
 ? TITLE OF INVENTION: CANCER  
 ? NUMBER OF SEQUENCES: 24  
 ? CORRESPONDENCE ADDRESS:  
 ?

ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/425,061  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: OSMAN, Richard A  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: A-59563-3/DJB/RAO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 494-8700  
TELEFAX: (415) 494-8771  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5709 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cdna  
US-08-425-061-8

Query Match 73.6%; Score 16.2; DB 2; Length 5709;  
Best Local Similarity 85.7%; Pred. NO. 2e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTAACCTGGA 22  
||||| ||||| ||||| ||||| |||||  
Db 5439 CATGCCACAGATCAACTGGA 5459

RESULT 26  
US-08-425-061-9  
Sequence 9, Application US/08425061  
Patent No. 5622829  
GENERAL INFORMATION:  
APPLICANT: KING, Mary-Claire  
APPLICANT: FRIEDMAN, Lori  
APPLICANT: OSTERMEYER, Beth  
APPLICANT: ROWELL, Sarah  
APPLICANT: LYNCH, Eric  
APPLICANT: SZABO, Csilla  
APPLICANT: LEE, Ming  
TITLE OF INVENTION: GENETIC MARKERS FOR BREAST AND OVARIAN  
TITLE OF INVENTION: CANCER  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/425,061  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:

NAME: OSMAN, Richard A  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: A-59563-3/DJB/RAO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 494-8700  
TELEFAX: (415) 494-8771  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5709 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cdna  
US-08-425-061-9

Query Match 73.6%; Score 16.2; DB 2; Length 5709;  
Best Local Similarity 85.7%; Pred. NO. 2e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTAACCTGGA 22  
||||| ||||| ||||| ||||| |||||  
Db 5439 CATGCCACAGATCAACTGGA 5459

RESULT 27  
US-08-825-886-2  
Sequence 2, Application US/08825886  
Patent No. 5821328  
GENERAL INFORMATION:  
APPLICANT: KING, Mary-Claire  
APPLICANT: FRIEDMAN, Lori  
APPLICANT: OSTERMEYER, Beth  
APPLICANT: ROWELL, Sarah  
APPLICANT: LYNCH, Eric  
APPLICANT: SZABO, Csilla  
APPLICANT: LEE, Ming  
TITLE OF INVENTION: GENETIC MARKERS FOR BREAST AND OVARIAN  
TITLE OF INVENTION: CANCER  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/825,886  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/425,061  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: OSMAN, Richard A  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: A-59563-3/DJB/RAO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 494-8700  
TELEFAX: (415) 494-8771  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5709 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear

```
; MOLECULE TYPE: cDNA
US-08-825-886-2

Query Match      73.6%; Score 16.2; DB 2; Length 5709;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      2 CATGCTCACACATTAACCTGGA 22
      ||||| ||||| ||||| |||||
Db      5439 CATGCCACAGATCACTGGA 5459

RESULT 28
US-08-825-886-7
; Sequence 7, Application US/08825886
; Patent No. 5821328
; GENERAL INFORMATION:
; APPLICANT: KING, Mary-Claire
; APPLICANT: FRIEDMAN, Lori
; APPLICANT: OSTERMEYER, Beth
; APPLICANT: ROWELL, Sarah
; APPLICANT: LYNCH, Eric
; APPLICANT: SZABO, Csilla
; APPLICANT: LEE, Ming
; TITLE OF INVENTION: GENETIC MARKERS FOR BREAST AND OVARIAN
; TITLE OF INVENTION: CANCER
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/825,886
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/425,061
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-59563-3/DJB/RAO
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5709 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-825-886-7

Query Match      73.6%; Score 16.2; DB 2; Length 5709;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      2 CATGCTCACACATTAACCTGGA 22
      ||||| ||||| ||||| |||||
Db      5439 CATGCCACAGATCACTGGA 5459

RESULT 29
US-08-825-886-7
; Sequence 7, Application US/08825886
; Patent No. 5821328
; GENERAL INFORMATION:
; APPLICANT: KING, Mary-Claire
; APPLICANT: FRIEDMAN, Lori
; APPLICANT: OSTERMEYER, Beth
; APPLICANT: ROWELL, Sarah
; APPLICANT: LYNCH, Eric
; APPLICANT: SZABO, Csilla
; APPLICANT: LEE, Ming
; TITLE OF INVENTION: GENETIC MARKERS FOR BREAST AND OVARIAN
; TITLE OF INVENTION: CANCER
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/825,886
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/425,061
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-59563-3/DJB/RAO
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5709 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-825-886-7

Query Match      73.6%; Score 16.2; DB 2; Length 5709;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      2 CATGCTCACACATTAACCTGGA 22
      ||||| ||||| ||||| |||||
Db      5439 CATGCCACAGATCACTGGA 5459

RESULT 30
US-08-825-886-9
; Sequence 9, Application US/08825886
; Patent No. 5821328
; GENERAL INFORMATION:
; APPLICANT: KING, Mary-Claire
; APPLICANT: FRIEDMAN, Lori
; APPLICANT: OSTERMEYER, Beth
; APPLICANT: ROWELL, Sarah
; APPLICANT: LYNCH, Eric
; APPLICANT: SZABO, Csilla
; APPLICANT: LEE, Ming
; TITLE OF INVENTION: GENETIC MARKERS FOR BREAST AND OVARIAN
; TITLE OF INVENTION: CANCER
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/825,886
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/425,061
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-59563-3/DJB/RAO
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5709 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-825-886-8

Query Match      73.6%; Score 16.2; DB 2; Length 5709;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      2 CATGCTCACACATTAACCTGGA 22
      ||||| ||||| ||||| |||||
Db      5439 CATGCCACAGATCACTGGA 5459
```

```
/
/ NUMBER OF SEQUENCES: 24
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
/ STREET: 4 Embarcadero Center, Suite 3400
/ CITY: San Francisco
/ STATE: California
/ COUNTRY: USA
/ ZIP: 94111-4187
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/825,886
/ FILING DATE:
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/425,061
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: OSMAN, Richard A
/ REGISTRATION NUMBER: 36,627
/ REFERENCE/DOCKET NUMBER: A-59563-3/DJB/RAO
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (415) 494-8700
/ TELEFAX: (415) 494-8771
/ TELEX: 910 277299
/ INFORMATION FOR SEQ ID NO: 9:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 5709 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: cdna
/ US-08-825-886-9

Query Match 73.6%; Score 16.2; DB 2; Length 5709;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTAAGTGA 22
||||| ||||| ||||| ||||| |||||
Db 5439 CATGCCACAGATCAACTGGA 5459

RESULT 31
US-08-899-890-2
/ Sequence 2, Application US/08989890
/ Patent No. 6512091
/ GENERAL INFORMATION:
/ APPLICANT: KING, Mary-Claire
/ FRIEDMAN, Lori
/ OSTERMEYER, Beth
/ ROWELL, Sarah
/ LYNCH, Eric
/ SZABO, Csilla
/ LEE, Ming
/ TITLE OF INVENTION: GENETIC MARKERS FOR BREAST AND OVARIAN
/ CANCER
/ NUMBER OF SEQUENCES: 24
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
/ STREET: 4 Embarcadero Center, Suite 3400
/ CITY: San Francisco
/ STATE: California
/ COUNTRY: USA
/ ZIP: 94111-4187
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/825,886
/ FILING DATE:
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/425,061
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: OSMAN, Richard A
/ REGISTRATION NUMBER: 36,627
/ REFERENCE/DOCKET NUMBER: A-59563-3/DJB/RAO
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (415) 494-8700
/ TELEFAX: (415) 494-8771
/ TELEX: 910 277299
/ INFORMATION FOR SEQ ID NO: 9:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 5709 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: cdna
/ US-08-825-886-9
```

```
/
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/989,890
/ FILING DATE: 12-Dec-1997
/ CLASSIFICATION: <Unknown>
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US/08/825,886
/ FILING DATE: <Unknown>
/ APPLICATION NUMBER: 08/425,061
/ FILING DATE: <Unknown>
/ ATTORNEY/AGENT INFORMATION:
/ NAME: OSMAN, Richard A
/ REGISTRATION NUMBER: 36,627
/ REFERENCE/DOCKET NUMBER: A-59563-3/DJB/RAO
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (415) 494-8700
/ TELEFAX: (415) 494-8771
/ TELEX: 910 277299
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 5709 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: cdna
/ SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-08-989-890-2

Query Match 73.6%; Score 16.2; DB 3; Length 5709;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTAAGTGA 22
||||| ||||| ||||| ||||| |||||
Db 5439 CATGCCACAGATCAACTGGA 5459

RESULT 32
US-08-989-890-7
/ Sequence 7, Application US/08989890
/ Patent No. 6512091
/ GENERAL INFORMATION:
/ APPLICANT: KING, Mary-Claire
/ FRIEDMAN, Lori
/ OSTERMEYER, Beth
/ ROWELL, Sarah
/ LYNCH, Eric
/ SZABO, Csilla
/ LEE, Ming
/ TITLE OF INVENTION: GENETIC MARKERS FOR BREAST AND OVARIAN
/ CANCER
/ NUMBER OF SEQUENCES: 24
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
/ STREET: 4 Embarcadero Center, Suite 3400
/ CITY: San Francisco
/ STATE: California
/ COUNTRY: USA
/ ZIP: 94111-4187
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/989,890
/ FILING DATE: 12-Dec-1997
/ CLASSIFICATION: <Unknown>
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US/08/825,886
/ FILING DATE: <Unknown>
/ APPLICATION NUMBER: 08/425,061
/ FILING DATE: <Unknown>
/ ATTORNEY/AGENT INFORMATION:
```



```
;
; NAME: OSMAN, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-59563-3/DJB/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299
;
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5709 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-08-989-890-7

Query Match 73.6%; Score 16.2; DB 3; Length 5709;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTAACCTGGA 22
DB 5439 CATGCCACAGATCAACTGGA 5459

RESULT 33
US-08-989-890-8
; Sequence 8, Application US/08989890
; Patent No. 6512091
; GENERAL INFORMATION:
; APPLICANT: KING, Mary-Claire
; FRIEDMAN, Lori
; OSTERMEYER, Beth
; ROWELL, Sarah
; LYNCH, Eric
; SZABO, Csilla
; LEE, Ming
;
; TITLE OF INVENTION: GENETIC MARKERS FOR BREAST AND OVARIAN
; CANCER
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 12-Dec-1997
; APPLICATION NUMBER: US/08/989,890
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/825,886
; FILING DATE: <Unknown>
; APPLICATION NUMBER: 08/425,061
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-59563-3/DJB/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299
;
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5709 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-08-989-890-8

Query Match 73.6%; Score 16.2; DB 3; Length 5709;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
;
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-08-989-890-8

Query Match 73.6%; Score 16.2; DB 3; Length 5709;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTAACCTGGA 22
DB 5439 CATGCCACAGATCAACTGGA 5459

RESULT 34
US-08-989-890-9
; Sequence 9, Application US/08989890
; Patent No. 6512091
; GENERAL INFORMATION:
; APPLICANT: KING, Mary-Claire
; FRIEDMAN, Lori
; OSTERMEYER, Beth
; ROWELL, Sarah
; LYNCH, Eric
; SZABO, Csilla
; LEE, Ming
;
; TITLE OF INVENTION: GENETIC MARKERS FOR BREAST AND OVARIAN
; CANCER
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 12-Dec-1997
; APPLICATION NUMBER: US/08/989,890
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/825,886
; FILING DATE: <Unknown>
; APPLICATION NUMBER: 08/425,061
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-59563-3/DJB/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299
;
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5709 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-08-989-890-9

Query Match 73.6%; Score 16.2; DB 3; Length 5709;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      2 CATGCTCACACATTAAGTGA 22
      ||||| ||||| ||||| |||||
Db      5439 CATGCCACAGATCAACTGGA 5459

RESULT 35
US-08-425-061-6
; Sequence 6, Application US/08425061
; Patent No. 5622829
; GENERAL INFORMATION:
; APPLICANT: KING, Mary-Claire
; APPLICANT: FRIEDMAN, Lori
; APPLICANT: OSTERMEYER, Beth
; APPLICANT: ROWELL, Sarah
; APPLICANT: LYNCH, Eric
; APPLICANT: SZABO, Csilla
; APPLICANT: LEE, Ming
; TITLE OF INVENTION: GENETIC MARKERS FOR BREAST AND OVARIAN
; TITLE OF INVENTION: CANCER
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/425,061
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/425,061
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-59563-3/DJB/RAO
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5710 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-825-886-6

Query Match      73.6%; Score 16.2; DB 2; Length 5710;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 CATGCTCACACATTAAGTGA 22
      ||||| ||||| ||||| |||||
Db      5440 CATGCCACAGATCAACTGGA 5460

RESULT 37
US-08-989-890-6
; Sequence 6, Application US/08989890
; Patent No. 6512091
; GENERAL INFORMATION:
; APPLICANT: KING, Mary-Claire
; APPLICANT: FRIEDMAN, Lori
; APPLICANT: OSTERMEYER, Beth
; APPLICANT: ROWELL, Sarah
; APPLICANT: LYNCH, Eric
; APPLICANT: SZABO, Csilla
; APPLICANT: LEE, Ming
; TITLE OF INVENTION: GENETIC MARKERS FOR BREAST AND OVARIAN
; TITLE OF INVENTION: CANCER
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
```

;/ MEDIUM TYPE: Floppy disk  
;/ COMPUTER: IBM PC compatible  
;/ OPERATING SYSTEM: PC-DOS/MS-DOS  
;/ SOFTWARE: Patent In Release #1.0, Version #1.30  
;/ CURRENT APPLICATION DATA:  
;/ APPLICATION NUMBER: US/08/989,890  
;/ FILING DATE: 12-Dec-1997  
;/ CLASSIFICATION: <Unknown>  
;/ PRIOR APPLICATION DATA:  
;/ APPLICATION NUMBER: US/08/825,886  
;/ FILING DATE: <Unknown>  
;/ APPLICATION NUMBER: 08/425,061  
;/ FILING DATE: <Unknown>  
;/ ATTORNEY/AGENT INFORMATION:  
;/ NAME: OSMAN, Richard A  
;/ REGISTRATION NUMBER: 36,627  
;/ REFERENCE/DOCKET NUMBER: A-59563-3/DJB/RAO  
;/ TELECOMMUNICATION INFORMATION:  
;/ TELEPHONE: (415) 494-8700  
;/ TELEX: 910 277299  
;/ INFORMATION FOR SEQ ID NO: 6:  
;/ SEQUENCE CHARACTERISTICS:  
;/ LENGTH: 5710 base pairs  
;/ TYPE: nucleic acid  
;/ STRANDEDNESS: double  
;/ TOPOLOGY: linear  
;/ MOLECULE TYPE: cDNA  
;/ SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
US-08-989-890-6  
  
Query Match 73.6%; Score 16.2; DB 3; Length 5710;  
Best Local Similarity 85.7%; Pred. No. 2e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 2 CATGCTCACACATTAACTGGA 22  
Db 5440 CATGCCACAGATCAACTGGA 5460  
  
RESULT 38  
US-08-425-061-4  
;/ Sequence 4, Application US/08425061  
;/ Patent No. 5622829  
;/ GENERAL INFORMATION:  
;/ APPLICANT: KING, Mary-Claire  
;/ APPLICANT: FRIEDMAN, Lori  
;/ APPLICANT: OSTERMEYER, Beth  
;/ APPLICANT: ROWELL, Sarah  
;/ APPLICANT: LYNCH, Eric  
;/ APPLICANT: SZABO, Csilla  
;/ APPLICANT: LEE, Ming  
;/ TITLE OF INVENTION: GENETIC MARKERS FOR BREAST AND OVARIAN  
;/ NUMBER OF SEQUENCES: 24  
;/ CORRESPONDENCE ADDRESS:  
;/ ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT  
;/ STREET: 4 Embarcadero Center, Suite 3400  
;/ CITY: San Francisco  
;/ STATE: California  
;/ COUNTRY: USA  
;/ ZIP: 94111-4187  
;/ COMPUTER READABLE FORM:  
;/ MEDIUM TYPE: Floppy disk  
;/ COMPUTER: IBM PC compatible  
;/ OPERATING SYSTEM: PC-DOS/MS-DOS  
;/ SOFTWARE: Patent In Release #1.0, Version #1.30  
;/ CURRENT APPLICATION DATA:  
;/ APPLICATION NUMBER: US/08/425,061  
;/ FILING DATE:  
;/ CLASSIFICATION: 435  
;/ ATTORNEY/AGENT INFORMATION:  
;/ NAME: OSMAN, Richard A  
;/ REGISTRATION NUMBER: 36,627  
;/ REFERENCE/DOCKET NUMBER: A-59563-3/DJB/RAO  
;/ TELECOMMUNICATION INFORMATION:  
;/ TELEPHONE: (415) 494-8700  
;/ TELEX: 910 277299  
;/ INFORMATION FOR SEQ ID NO: 10:  
;/ SEQUENCE CHARACTERISTICS:  
;/ LENGTH: 5711 base pairs  
;/ TYPE: nucleic acid  
;/ STRANDEDNESS: double  
;/ TOPOLOGY: linear  
;/ MOLECULE TYPE: cDNA  
;/ US-08-425-061-10

;/ REGISTRATION NUMBER: 36,627  
;/ REFERENCE/DOCKET NUMBER: A-59563-3/DJB/RAO  
;/ TELECOMMUNICATION INFORMATION:  
;/ TELEPHONE: (415) 494-8700  
;/ TELEFAX: (415) 494-8771  
;/ TELEX: 910 277299  
;/ INFORMATION FOR SEQ ID NO: 4:  
;/ SEQUENCE CHARACTERISTICS:  
;/ LENGTH: 5711 base pairs  
;/ TYPE: nucleic acid  
;/ STRANDEDNESS: double  
;/ TOPOLOGY: linear  
;/ MOLECULE TYPE: cDNA  
;/ US-08-425-061-4  
  
Query Match 73.6%; Score 16.2; DB 2; Length 5711;  
Best Local Similarity 85.7%; Pred. No. 2e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 2 CATGCTCACACATTAACTGGA 22  
Db 5441 CATGCCACAGATCAACTGGA 5461  
  
RESULT 39  
US-08-425-061-10  
;/ Sequence 10, Application US/08425061  
;/ Patent No. 5622829  
;/ GENERAL INFORMATION:  
;/ APPLICANT: KING, Mary-Claire  
;/ APPLICANT: FRIEDMAN, Lori  
;/ APPLICANT: OSTERMEYER, Beth  
;/ APPLICANT: ROWELL, Sarah  
;/ APPLICANT: LYNCH, Eric  
;/ APPLICANT: SZABO, Csilla  
;/ APPLICANT: LEE, Ming  
;/ TITLE OF INVENTION: GENETIC MARKERS FOR BREAST AND OVARIAN  
;/ NUMBER OF SEQUENCES: 24  
;/ CORRESPONDENCE ADDRESS:  
;/ ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT  
;/ STREET: 4 Embarcadero Center, Suite 3400  
;/ CITY: San Francisco  
;/ STATE: California  
;/ COUNTRY: USA  
;/ ZIP: 94111-4187  
;/ COMPUTER READABLE FORM:  
;/ MEDIUM TYPE: Floppy disk  
;/ COMPUTER: IBM PC compatible  
;/ OPERATING SYSTEM: PC-DOS/MS-DOS  
;/ SOFTWARE: Patent In Release #1.0, Version #1.30  
;/ CURRENT APPLICATION DATA:  
;/ APPLICATION NUMBER: US/08/425,061  
;/ FILING DATE:  
;/ CLASSIFICATION: 435  
;/ ATTORNEY/AGENT INFORMATION:  
;/ NAME: OSMAN, Richard A  
;/ REGISTRATION NUMBER: 36,627  
;/ REFERENCE/DOCKET NUMBER: A-59563-3/DJB/RAO  
;/ TELECOMMUNICATION INFORMATION:  
;/ TELEPHONE: (415) 494-8700  
;/ TELEFAX: (415) 494-8771  
;/ TELEX: 910 277299  
;/ INFORMATION FOR SEQ ID NO: 10:  
;/ SEQUENCE CHARACTERISTICS:  
;/ LENGTH: 5711 base pairs  
;/ TYPE: nucleic acid  
;/ STRANDEDNESS: double  
;/ TOPOLOGY: linear  
;/ MOLECULE TYPE: cDNA  
;/ US-08-425-061-10  
  
Query Match 73.6%; Score 16.2; DB 2; Length 5711;

Best Local Similarity 85.7%; Pred. No. 2e+02; Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTAACTGGA 22  
||||| ||||| ||||| ||||| |||||  
Db 5441 CATGCCACACAGATCAACTGGA 5461

RESULT 40  
US-08-598-591-1  
; Sequence 1, Application US/08598591  
; Patent No. 5654155  
; GENERAL INFORMATION:  
; APPLICANT: Allen, Antonette C.  
; APPLICANT: Alvares, Christopher P.  
; APPLICANT: Critz, Brenda S.  
; APPLICANT: Murphy Patricia D.  
; APPLICANT: Olson, Sheri J.  
; APPLICANT: Schelter, Denise B.  
; APPLICANT: Zeng, Bin  
; TITLE OF INVENTION: A Consensus Sequence of the Human BRCA1 Gene  
; Patent No. 5654155  
; NUMBER OF SEQUENCES: 74  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
; STREET: 699 Prince St.  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22314  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/598,591  
; FILING DATE: herewith  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Swecker, Robert S.  
; REGISTRATION NUMBER: 19,885  
; REFERENCE/DOCKET NUMBER: 020160-282  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-836-6620  
; TELEFAX: 703-836-2021  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5711 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; STRAIN: BRCA1  
; POSITION IN GENOME:  
; CHROMOSOME/SEGMENT: 17  
; MAP POSITION: 17q21  
US-08-598-591-1

Query Match 73.6%; Score 16.2; DB 2; Length 5711;  
Best Local Similarity 85.7%; Pred. No. 2e+02; Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTAACTGGA 22  
||||| ||||| ||||| ||||| |||||  
Db 5441 CATGCCACACAGATCAACTGGA 5461

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2005, 18:34:25 ; Search time 662.716 Seconds  
(without alignments)  
274.516 Million cell updates/sec

Title: US-10-627-757-28

Perfect score: 22  
Sequence: 1 tcattgctcacacattactgga 22

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 413469005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_NA\_Main:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US11\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	100.0	22	7	US-10-627-757-28
2	22	100.0	1116	7	US-10-627-757-8
3	22	100.0	46951	6	US-10-091-281-2
4	19.4	88.2	577	4	US-09-925-065A-194829
5	19.4	88.2	731	6	US-10-029-386-22960
6	19.4	88.2	4007	8	US-10-852-943-95
7	19.4	88.2	4557	9	US-10-450-763-28227
8	17.8	80.9	556	4	US-09-925-065A-369426
9	17.8	80.9	678	5	US-10-027-632-238976
10	17.8	80.9	678	6	US-10-027-632-238976
11	17.8	80.9	842	7	US-10-425-114-32603
12	17.8	80.9	938	8	US-10-425-115-48809
13	17.2	78.2	318	7	US-10-424-599-43740
14	17.2	78.2	600	9	US-10-972-079-34339
15	17.2	78.2	782	7	US-10-437-963-86498
16	16.8	76.4	25	7	US-10-719-956-89691
17	16.8	76.4	466	5	US-10-027-632-180356
18	16.8	76.4	466	5	US-10-027-632-180356
19	16.8	76.4	466	6	US-10-027-632-180356
20	16.8	76.4	466	6	US-10-027-632-180356
21	16.8	76.4	520	5	US-10-027-632-164946
22	16.8	76.4	520	6	US-10-027-632-164946
23	16.8	76.4	534	4	US-09-925-065A-14322

24	16.8	76.4	552	4	US-09-925-065A-844974	Sequence 844974,
25	16.8	76.4	600	9	US-10-972-079-2711	Sequence 2711, Ap
26	16.8	76.4	600	9	US-10-972-079-2712	Sequence 2712, Ap
27	16.8	76.4	600	9	US-10-972-079-2713	Sequence 2713, Ap
28	16.8	76.4	622	4	US-09-925-065A-319331	Sequence 319331,
29	16.8	76.4	625	4	US-09-925-065A-579357	Sequence 579357,
30	16.8	76.4	875	4	US-09-925-065A-554033	Sequence 554033,
31	16.8	76.4	875	4	US-09-925-065A-554034	Sequence 554034,
32	16.8	76.4	1068	5	US-10-027-632-117652	Sequence 117652,
33	16.8	76.4	1068	5	US-10-027-632-117653	Sequence 117653,
34	16.8	76.4	1068	6	US-10-027-632-117652	Sequence 117652,
35	16.8	76.4	1068	6	US-10-027-632-117653	Sequence 117653,
36	16.8	76.4	156318	8	US-10-741-600-115754	Sequence 17574, A
37	16.8	76.4	401616	5	US-10-087-192-1630	Sequence 1630, Ap
38	16.8	76.4	599001	9	US-10-317-869A-4	Sequence 4, Appli
39	16.4	74.5	613	4	US-09-925-065A-496606	Sequence 496606,
40	16.4	74.5	788	8	US-10-425-115-180020	Sequence 180020,
41	16.4	74.5	1282	7	US-10-424-599-80379	Sequence 80379, A
42	16.4	74.5	15845	7	US-10-668-767-3	Sequence 3, Appli
43	16.4	74.5	247544	7	US-10-322-696-55	Sequence 55, Appl
44	16.2	73.6	25	7	US-10-719-956-580893	Sequence 580893,
45	16.2	73.6	65	6	US-10-388-360-69	Sequence 69, Appl

ALIGNMENTS

RESULT 1

US-10-627-757-28  
; Sequence 28, Application US/10627757  
; Publication No. US20040091914A1  
; GENERAL INFORMATION:  
; APPLICANT: KOUCHI YASUHIRO  
; APPLICANT: MASASGO AKINORI  
; APPLICANT: TAKAHATI TAKAYUKI  
; TITLE OF INVENTION: GENE ASSAY METHOD FOR PREDICTING GLAUCOMA ONSET RISK  
; FILE REFERENCE: Q76319  
; CURRENT APPLICATION NUMBER: US/10/627,757  
; CURRENT FILING DATE: 2003-07-28  
; PRIOR APPLICATION NUMBER: JP P2002-226612  
; PRIOR FILING DATE: 2002-08-02  
; NUMBER OF SEQ ID NOS: 40  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 28  
; LENGTH: 22  
; TYPE: DNA  
; ORGANISM: Artificial  
; FEATURE:  
; OTHER INFORMATION: Designed DNA based on OPTN gene  
US-10-627-757-28

Query Match 100.0%; Score 22; DB 7; Length 22;  
Best Local Similarity 100.0%; Pred.No. 0.81; 0; Indels 0; Gaps 0;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCATGCTCACACATTAACTGGA 22

Db 1 TCATGCTCACACATTAACTGGA 22

RESULT 2

US-10-627-757-8/c  
; Sequence 8, Application US/10627757  
; Publication No. US20040091914A1  
; GENERAL INFORMATION:  
; APPLICANT: KOUCHI YASUHIRO  
; APPLICANT: MASASGO AKINORI  
; APPLICANT: TAKAHATI TAKAYUKI  
; TITLE OF INVENTION: GENE ASSAY METHOD FOR PREDICTING GLAUCOMA ONSET RISK  
; FILE REFERENCE: Q76319  
; CURRENT APPLICATION NUMBER: US/10/627,757  
; CURRENT FILING DATE: 2003-07-28  
; PRIOR APPLICATION NUMBER: JP P2002-226612

```
; PRIOR FILING DATE: 2002-08-02
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 1116
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-627-757-8

Query Match      100.0%; Score 22; DB 7; Length 1116;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 TCATGCTCACATTAACTGGA 22
Db      786 TCATGCTCACATTAACTGGA 765

RESULT 3
US-10-091-281-2/c
; Sequence 2, Application US/10091281
; Publication No. US20030190617A1
; GENERAL INFORMATION:
; APPLICANT: RAYMOND, VINCENT
; APPLICANT: SI, ERWIN
; APPLICANT: MORISSETTE, JEAN
; TITLE OF INVENTION: OPTINEURIN NUCLEIC ACID MOLECULES AND USES THEREOF
; FILE REFERENCE: 13587.338
; CURRENT APPLICATION NUMBER: US/10/091,281
; CURRENT FILING DATE: 2002-03-06
; NUMBER OF SEQ ID NOS: 463
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 46951
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 391
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 691
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 709
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 887
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 894
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 987
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 1112
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 1505
; OTHER INFORMATION: insertion of additional c residue
; NAME/KEY: allele
; LOCATION: 1606
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 2405
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 2606
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 3313
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
```

```
; NAME/KEY: allele
; LOCATION: 3555
; OTHER INFORMATION: insertion of additional t residue
; NAME/KEY: allele
; LOCATION: 3625
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 3629
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 3882
; OTHER INFORMATION: insertion of additional t residue
; NAME/KEY: allele
; LOCATION: 3988
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: allele
; LOCATION: 4452
; OTHER INFORMATION: single nucleotide polymorphism (SNP)
; NAME/KEY: repeat region
; LOCATION: 598..878
; OTHER INFORMATION: repeat element
; NAME/KEY: repeat region
; LOCATION: 938..957
; OTHER INFORMATION: Short repeat element
; NAME/KEY: repeat region
; LOCATION: 1002..1329
; OTHER INFORMATION: ALU repeat element
; NAME/KEY: repeat region
; LOCATION: 2288..2587
; OTHER INFORMATION: ALU repeat element
; NAME/KEY: misc_feature
; LOCATION: 5054
; OTHER INFORMATION: putative transcription start site
US-10-091-281-2

Query Match      100.0%; Score 22; DB 6; Length 46951;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 TCATGCTCACATTAACTGGA 22
Db      29124 TCATGCTCACATTAACTGGA 29103

RESULT 4
US-09-925-065A-194829
; Sequence 194829, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 194829
; LENGTH: 577
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-194829
```

```
Query Match      88.2%; Score 19.4; DB 4; Length 577;
Best Local Similarity 95.2%; Pred. No. 24;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTAACCTGGA 22
    ||||| ||||| ||||| ||||| |||||
Db 8 CATGCTTACACATTAACCTGGA 28

RESULT 5
US-10-029-386-22960/c
; Sequence 22960, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Hanzel, David R.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 22960
; LENGTH: 731
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL121948.7
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.46
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.4
; OTHER INFORMATION: SWISSPROT HIT: P81133, EVALUATION 0.00e+00
; OTHER INFORMATION: NT HIT: g113631082, EVALUATION 0.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: BF679225.1, EVALUATION 1.00e+00
US-10-029-386-22960

Query Match      88.2%; Score 19.4; DB 6; Length 731;
Best Local Similarity 95.2%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTAACCTGGA 22
    ||||| ||||| ||||| ||||| |||||
Db 177 CATGCTTACACATTAACCTGGA 157

RESULT 6
US-10-852-943-95
; Sequence 95, Application US/10852943
; Publication No. US20050037388A1
; GENERAL INFORMATION:
; APPLICANT: University of Geneva
; APPLICANT: Stylianos, Antonarakis
; APPLICANT: Deutsch, Samuel
; TITLE OF INVENTION: METHOD FOR DETECTING DISEASES CAUSED BY CHROMOSOMAL IMBALANCES
; FILE REFERENCE: 27067/2005
; CURRENT APPLICATION NUMBER: US/10/852,943
; CURRENT FILING DATE: 2004-05-25
; PRIOR APPLICATION NUMBER: US 60/300,266
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 10/177,063
; PRIOR FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 95
; LENGTH: 4007
; TYPE: DNA
; ORGANISM: Homo sapiens
```

```
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (125)..(125)
; OTHER INFORMATION: n is a, c, g, or t
US-10-852-943-95

Query Match      88.2%; Score 19.4; DB 8; Length 4007;
Best Local Similarity 95.2%; Pred. No. 31;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTAACCTGGA 22
    ||||| ||||| ||||| ||||| |||||
Db 2341 CATGCTTACACATTAACCTGGA 2361

RESULT 7
US-10-450-763-28227
; Sequence 28227, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 28227
; LENGTH: 4557
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIMILAR
; LOCATION: (767)..(3064)
; OTHER INFORMATION: 100% homologous to Homo sapiens hSIM1, accession number
; OTHER INFORMATION: U70212, Smith-Waterman Score=4081.
US-10-450-763-28227

Query Match      88.2%; Score 19.4; DB 9; Length 4557;
Best Local Similarity 95.2%; Pred. No. 32;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTAACCTGGA 22
    ||||| ||||| ||||| ||||| |||||
Db 2891 CATGCTTACACATTAACCTGGA 2911

RESULT 8
US-09-925-065A-369426/c
; Sequence 369426, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
```

```
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 369426
; LENGTH: 596
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-369426

Query Match      80.9%; Score 17.8; DB 4; Length 596;
Best Local Similarity 90.5%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCATGCTCACACATTAACCTGG 21
    ||||| ||||| ||||| |||||
DB 413 TCATGCAGACACATTAACCTGG 393

RESULT 9
US-10-027-632-238976
; Sequence 238976, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 238976
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-238976

Query Match      80.9%; Score 17.8; DB 5; Length 678;
Best Local Similarity 90.5%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCATGCTCACACATTAACCTGG 21
    ||||| ||||| ||||| |||||
DB 540 TCATGCAGACACATTAACCTGG 560

RESULT 10
US-10-027-632-238976
; Sequence 238976, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
```

```
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 238976
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-238976

Query Match      80.9%; Score 17.8; DB 6; Length 678;
Best Local Similarity 90.5%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCATGCTCACACATTAACCTGG 21
    ||||| ||||| ||||| |||||
DB 540 TCATGCAGACACATTAACCTGG 560

RESULT 11
US-10-425-114-32603
; Sequence 32603, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 32603
; LENGTH: 842
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLM017009C07_FLI
US-10-425-114-32603

Query Match      80.9%; Score 17.8; DB 7; Length 842;
Best Local Similarity 90.5%; Pred. No. 1.6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTAACCTGGA 22
    ||||| ||||| ||||| |||||
DB 49 CATGAGCACACATTAACCTGGA 69

RESULT 12
US-10-425-115-48809
; Sequence 48809, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
```



```
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 48809
; LENGTH: 938
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_14451C.1
US-10-425-115-48809

Query Match      80.9%; Score 17.8; DB 8; Length 938;
Best Local Similarity 90.5%; Pred. No. 1.6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY  2 CATGCTCACACATTAACTGGA 22
Db  81 CATGAGCACACATTAACTGGA 101

RESULT 13
US-10-424-599-43740
; Sequence 43740, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 43740
; LENGTH: 318
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_139497C.1
US-10-424-599-43740

Query Match      78.2%; Score 17.2; DB 7; Length 318;
Best Local Similarity 86.4%; Pred. No. 2.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY  1 TCATGCTCACACATTAACTGGA 22
Db  188 TAATGCTCACACATTACATGGA 209

RESULT 14
US-10-972-079-34339/c
; Sequence 34339, Application US/10972079
; Publication No. US20050153317A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: ROSENFELD, David
; APPLICANT: KERR, Richard
; APPLICANT: BATES, Stephen
; APPLICANT: HOLM, Ton
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEER
; TITLE OF INVENTION: LIVESTOCK
; FILE REFERENCE: MM1110-2
; CURRENT APPLICATION NUMBER: US/10/972,079
; CURRENT FILING DATE: 2004-10-22
; PRIOR APPLICATION NUMBER: US 60/514,333
; PRIOR FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 96631
```

```
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 34339
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Chicken 19866894259434_1
US-10-972-079-34339

Query Match      78.2%; Score 17.2; DB 9; Length 600;
Best Local Similarity 86.4%; Pred. No. 3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY  1 TCATGCTCACACATTAACTGGA 22
Db  49 TCATGCTCTCAAAATTAATGGA 28

RESULT 15
US-10-437-963-86498
; Sequence 86498, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 86498
; LENGTH: 782
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_85533C.1
US-10-437-963-86498

Query Match      78.2%; Score 17.2; DB 7; Length 782;
Best Local Similarity 86.4%; Pred. No. 3.1e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY  1 TCATGCTCACACATTAACTGGA 22
Db  518 TCATGCTCACACATAAAAGGA 539

RESULT 16
US-10-719-956-89691
; Sequence 89691, Application US/10719956
; Publication No. US20040146910A1
; GENERAL INFORMATION:
; APPLICANT: Yue Mei Zhou
; TITLE OF INVENTION: Methods of Genetic Analysis of Rat
; FILE REFERENCE: 3527.1
; CURRENT APPLICATION NUMBER: US/10/719,956
; CURRENT FILING DATE: 2003-11-20
; PRIOR APPLICATION NUMBER: 60/427,836
; PRIOR FILING DATE: 2002 11 20
; NUMBER OF SEQ ID NOS: 699466
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 89691
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-10-719-956-89691

Query Match      76.4%; Score 16.8; DB 7; Length 25;
```

Best Local Similarity 90.0%; Pred. No. 3.1e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTAAGTGG 21  
          |||||  
Db 4 CATGCTCACACATGTAAGTGG 23

RESULT 17  
US-10-027-632-180356  
; Sequence 180356, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 180356  
; LENGTH: 466  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-180356

Query Match 76.4%; Score 16.8; DB 5; Length 466;  
Best Local Similarity 90.0%; Pred. No. 4.6e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 ATGCTCACACATTAAGTGG 22  
          |||||  
Db 18 ATGCTCACTCAGTAAGTGG 37

RESULT 18  
US-10-027-632-180357  
; Sequence 180357, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28

; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 180357  
; LENGTH: 466  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-180357

Query Match 76.4%; Score 16.8; DB 5; Length 466;  
Best Local Similarity 90.0%; Pred. No. 4.6e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 ATGCTCACACATTAAGTGG 22  
          |||||  
Db 18 ATGCTCACTCAGTAAGTGG 37

RESULT 19  
US-10-027-632-180356  
; Sequence 180356, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 180356  
; LENGTH: 466  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-180356

Query Match 76.4%; Score 16.8; DB 6; Length 466;  
Best Local Similarity 90.0%; Pred. No. 4.6e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 ATGCTCACACATTAAGTGG 22  
          |||||  
Db 18 ATGCTCACTCAGTAAGTGG 37

RESULT 20  
US-10-027-632-180357  
; Sequence 180357, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006

```
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 180357
; LENGTH: 466
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-180357
```

```
Query Match 76.4%; Score 16.8; DB 6; Length 466;
Best Local Similarity 90.0%; Pred. No. 4.6e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 3 ATGCTCACACATTAACCTGGA 22
| | | | | | | | | | | | | | | |
Db 18 ATGCTCACTCAGTAACCTGGA 37
```

## RESULT 21

```
US-10-027-632-164946/c
; Sequence 164946, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
```

```
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 164946
; LENGTH: 520
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-164946
```

```
Query Match 76.4%; Score 16.8; DB 5; Length 520;
Best Local Similarity 90.0%; Pred. No. 4.6e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 2 CATGCTCACACATTAACCTG 21
| | | | | | | | | | | | | | | |
Db 219 CATGCTCACACACTAACAGG 200
```

## RESULT 22

```
US-10-027-632-164946/c
; Sequence 164946, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
```

```
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 164946
; LENGTH: 520
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-164946
```

```
Query Match 76.4%; Score 16.8; DB 6; Length 520;
Best Local Similarity 90.0%; Pred. No. 4.6e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 2 CATGCTCACACATTAACCTG 21
| | | | | | | | | | | | | | | |
Db 219 CATGCTCACACACTAACAGG 200
```

## RESULT 23

```
US-09-925-065A-14322
; Sequence 14322, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
```

```
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14322
; LENGTH: 534
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-925-065A-14322
```

```
Query Match 76.4%; Score 16.8; DB 4; Length 534;
Best Local Similarity 90.0%; Pred. No. 4.6e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

QY 3 ATGCTCACATTAACTGGA 22  
|||||  
Db 381 ATGCTCACATACACTGGA 400

RESULT 24  
US-09-925-065A-844974  
; Sequence 844974, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; PRIOR FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 844974  
; LENGTH: 552  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-844974

Query Match 76.4%; Score 16.8; DB 4; Length 552;  
Best Local Similarity 90.0%; Pred. No. 4.7e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 ATGCTCACATTAACTGGA 22  
|||||  
Db 307 ATGCTCCACACTAACTGGA 326

RESULT 25  
US-10-972-079-2711  
; Sequence 2711, Application US/10972079  
; Publication No. US20050153317A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: ROSENFELD, David  
; APPLICANT: KERR, Richard  
; APPLICANT: BATES, Stephen  
; APPLICANT: HOLM, Tom  
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEEF  
; FILE REFERENCE: MM1110-2  
; CURRENT APPLICATION NUMBER: US/10/972,079  
; CURRENT FILING DATE: 2004-10-22  
; PRIOR APPLICATION NUMBER: US 60/514,333  
; PRIOR FILING DATE: 2003-10-24  
; NUMBER OF SEQ ID NOS: 96631  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2711  
; LENGTH: 600  
; TYPE: DNA  
; ORGANISM: Chicken 19866894089902\_5  
US-10-972-079-2711

Query Match 76.4%; Score 16.8; DB 9; Length 600;  
Best Local Similarity 90.0%; Pred. No. 4.7e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCATGCTCACATTAACTG 20  
|||||  
Db 255 TCTTGTCACATTAACTG 274

RESULT 26  
US-10-972-079-2712  
; Sequence 2712, Application US/10972079  
; Publication No. US20050153317A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: ROSENFELD, David  
; APPLICANT: KERR, Richard  
; APPLICANT: BATES, Stephen  
; APPLICANT: HOLM, Tom  
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEEF  
; FILE REFERENCE: MM1110-2  
; CURRENT APPLICATION NUMBER: US/10/972,079  
; CURRENT FILING DATE: 2004-10-22  
; PRIOR APPLICATION NUMBER: US 60/514,333  
; PRIOR FILING DATE: 2003-10-24  
; NUMBER OF SEQ ID NOS: 96631  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2712  
; LENGTH: 600  
; TYPE: DNA  
; ORGANISM: Chicken 19866894089902\_6  
US-10-972-079-2712

Query Match 76.4%; Score 16.8; DB 9; Length 600;  
Best Local Similarity 90.0%; Pred. No. 4.7e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCATGCTCACATTAACTG 20  
|||||  
Db 240 TCTTGTCACATTAACTG 259

RESULT 27  
US-10-972-079-2713  
; Sequence 2713, Application US/10972079  
; Publication No. US20050153317A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: ROSENFELD, David  
; APPLICANT: KERR, Richard  
; APPLICANT: BATES, Stephen  
; APPLICANT: HOLM, Tom  
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEEF  
; FILE REFERENCE: MM1110-2  
; CURRENT APPLICATION NUMBER: US/10/972,079  
; CURRENT FILING DATE: 2004-10-22  
; PRIOR APPLICATION NUMBER: US 60/514,333  
; PRIOR FILING DATE: 2003-10-24  
; NUMBER OF SEQ ID NOS: 96631  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2713  
; LENGTH: 600  
; TYPE: DNA  
; ORGANISM: Chicken 19866894089902\_7  
US-10-972-079-2713

Query Match 76.4%; Score 16.8; DB 9; Length 600;  
Best Local Similarity 90.0%; Pred. No. 4.7e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCATGCTCACATTAACTG 20  
|||||

```
Db      177 TCTTGCTGACACATTAACGTG 196

RESULT 28
US-09-925-065A-319331
; Sequence 319331, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 319331
; LENGTH: 622
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-319331

Query Match      76.4%; Score 16.8; DB 4; Length 622;
Best Local Similarity 90.0%; Pred. No. 4.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      3 ATGCTCACACATTAACGTGGA 22
      |||||
Db      379 ATGCTCACACATTAACGTGGA 398

RESULT 29
US-09-925-065A-579357
; Sequence 579357, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 579357
; LENGTH: 625
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-579357

Query Match      76.4%; Score 16.8; DB 4; Length 625;
Best Local Similarity 90.0%; Pred. No. 4.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db      177 TCTTGCTGACACATTAACGTG 196

RESULT 28
US-09-925-065A-319331
; Sequence 319331, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 319331
; LENGTH: 622
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-319331

Query Match      76.4%; Score 16.8; DB 4; Length 622;
Best Local Similarity 90.0%; Pred. No. 4.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      3 ATGCTCACACATTAACGTGGA 22
      |||||
Db      379 ATGCTCACACATTAACGTGGA 398

RESULT 29
US-09-925-065A-579357
; Sequence 579357, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 579357
; LENGTH: 625
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-579357

Query Match      76.4%; Score 16.8; DB 4; Length 625;
Best Local Similarity 90.0%; Pred. No. 4.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      3 ATGCTCACACATTAACGTGGA 22
      |||||
Db      114 ATGCTCACTCAGTAACTGGA 133

RESULT 31
US-09-925-065A-554034
; Sequence 554034, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 554034
; LENGTH: 875
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-554034

Query Match      76.4%; Score 16.8; DB 4; Length 875;
Best Local Similarity 90.0%; Pred. No. 5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

Query Match 76.4%; Score 16.8; DB 4; Length 875;  
Best Local Similarity 90.0%; Pred. No. 5e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 ATGCTCACACATTAACTGGA 22  
||||| ||| ||| ||| |||  
Db 114 ATGCTCACTCAGTAACTGGA 133

RESULT 32  
US-10-027-632-117652  
; Sequence 117652, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 117652  
; LENGTH: 1068  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-117652

Query Match 76.4%; Score 16.8; DB 5; Length 1068;  
Best Local Similarity 90.0%; Pred. No. 5.1e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCATGCTCACACATTAACTG 20  
||||| ||| ||| ||| |||  
Db 558 TCATGCACTCACATTAACTG 577

RESULT 33  
US-10-027-632-117653  
; Sequence 117653, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358

; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 117653  
; LENGTH: 1068  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-117653

Query Match 76.4%; Score 16.8; DB 5; Length 1068;  
Best Local Similarity 90.0%; Pred. No. 5.1e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCATGCTCACACATTAACTG 20  
||||| ||| ||| ||| |||  
Db 558 TCATGCACTCACATTAACTG 577

RESULT 34  
US-10-027-632-117652  
; Sequence 117652, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 117652  
; LENGTH: 1068  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-117652

Query Match 76.4%; Score 16.8; DB 6; Length 1068;  
Best Local Similarity 90.0%; Pred. No. 5.1e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCATGCTCACACATTAACTG 20  
||||| ||| ||| ||| |||  
Db 558 TCATGCACTCACATTAACTG 577

RESULT 35  
US-10-027-632-117653  
; Sequence 117653, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/218,006  
; CURRENT FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 117653  
; LENGTH: 1068  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-117653

Query Match 76.4%; Score 16.8; DB 6; Length 1068;  
Best Local Similarity 90.0%; Pred. No. 5.1e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCATGCTCACACATTAACGTG 20  
||||| |||||||  
Db 558 TCATGCACTCACATTAACGTG 577

RESULT 36  
US-10-741-600-17574/c  
; Sequence 17574, Application US/10741600  
; Publication No. US20050026169A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001499  
; CURRENT APPLICATION NUMBER: US/10/741,600  
; CURRENT FILING DATE: 2003-12-22  
; NUMBER OF SEQ ID NOS: 73997  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 17574  
; LENGTH: 156318  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; NAME/KEY: misc feature  
; LOCATION: (1)...(156318)  
; OTHER INFORMATION: n = A, T, C or G, or insertion/deletion polymorphism (see Tables 1-  
US-10-741-600-17574

Query Match 76.4%; Score 16.8; DB 8; Length 156318;  
Best Local Similarity 90.0%; Pred. No. 9.7e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCATGCTCACACATTAACGTG 20  
||||| |||||||  
Db 45219 TAATGCTTACACATTAACGTG 45200

RESULT 37  
US-10-087-192-1630/c  
; Sequence 1630, Application US/10087192  
; Publication No. US20020182586A1  
; GENERAL INFORMATION:  
; APPLICANT: Morris, David W.  
; APPLICANT: Engelhard, Eric K.  
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: CANCER  
; FILE REFERENCE: 529452000122

; CURRENT APPLICATION NUMBER: US/10/087,192  
; CURRENT FILING DATE: 2002-03-01  
; PRIOR APPLICATION NUMBER: US 09/747,377  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: US 09/798,586  
; PRIOR FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 2059  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1630  
; LENGTH: 401616  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)...(401616)  
; OTHER INFORMATION: n = A, T, C or G  
US-10-087-192-1630

Query Match 76.4%; Score 16.8; DB 5; Length 401616;  
Best Local Similarity 90.0%; Pred. No. 1.1e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 ATGCTCACACATTAACGTGA 22  
||||| |||||||  
Db 269113 ATGCTCCACACTAACTGGA 269094

RESULT 38  
US-10-317-869A-4/c  
; Sequence 4, Application US/10317869A  
; Publication No. US20050101000A1  
; GENERAL INFORMATION:  
; APPLICANT: Kenneth W. Dobie  
; APPLICANT: C. Frank Bennett  
; TITLE OF INVENTION: MODULATION OF PHOSPHODIESTERASE 4B EXPRESSION  
; FILE REFERENCE: RTS-0429  
; CURRENT APPLICATION NUMBER: US/10/317,869A  
; CURRENT FILING DATE: 2002-12-11  
; NUMBER OF SEQ ID NOS: 113  
; SEQ ID NO 4  
; LENGTH: 599001  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 55607-55706, 360227-360326, 363103-363202, 363428, 366093-366192,  
; LOCATION: 368528, 447532, 452653-452752, 457192, 457207-457306, 461808-461907,  
; LOCATION: 461921, 461928, 461942  
; OTHER INFORMATION: n = A, T, C or G  
US-10-317-869A-4

Query Match 76.4%; Score 16.8; DB 9; Length 599001;  
Best Local Similarity 90.0%; Pred. No. 1.1e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 ATGCTCACACATTAACGTGA 22  
||||| |||||||  
Db 474348 ATGCTCCACACTAACTGGA 474329

RESULT 39  
US-09-925-065A-496606  
; Sequence 496606, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 10827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24

```
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 496606
; LENGTH: 613
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-496606

Query Match          74.5%; Score 16.4; DB 4; Length 613;
Best Local Similarity 94.4%; Pred. No. 7.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      5  GCTCACACATTAACTGGA 22
         ||||| ||||| |||||
DB      387 GCTCACACAATAACTGGA 404

RESULT 40
US-10-425-115-180020
; Sequence 180020, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(5322)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 180020
; LENGTH: 788
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(788)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_95758C.1
US-10-425-115-180020

Query Match          74.5%; Score 16.4; DB 8; Length 788;
Best Local Similarity 94.4%; Pred. No. 7.7e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      4  TGCTCACACATTAACTGG 21
         ||||| ||||| |||||
DB      81  TGCTAACACATTAACTGG 98

Search completed: November 21, 2005, 05:03:45
Job time : 665.716 secs
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2005, 21:55:46 ; Search time 307.457 Seconds  
(without alignments)  
9.405 Million cell updates/sec

Title: US-10-627-757-28

Perfect score: 22  
Sequence: 1 tcatgctcacattactgga 22

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3196817 seqs, 65720914 residues

Total number of hits satisfying chosen parameters: 6393634

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_NA\_New.\*  
1: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq2.\*  
9: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq3.\*  
10: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	15	68.2	19	8	US-11-101-244-280543
2	15	68.2	19	9	US-11-083-784-280543
C 3	14.8	67.3	19	8	US-11-101-244-266902
4	14.8	67.3	19	8	US-11-101-244-1110410
C 5	14.8	67.3	19	9	US-11-083-784-266902
6	14.8	67.3	19	9	US-11-083-784-1110410
7	14.4	65.5	19	8	US-11-101-244-119918
C 8	14.4	65.5	19	8	US-11-101-244-993282
9	14.4	65.5	19	8	US-11-101-244-1048490
10	14.4	65.5	19	9	US-11-083-784-119918
C 11	14.4	65.5	19	9	US-11-083-784-993282
12	14.4	65.5	19	9	US-11-083-784-1048490
C 13	14.2	64.5	19	8	US-11-101-244-876983
C 14	14.2	64.5	19	9	US-11-083-784-876983
C 15	14.2	64.5	19	8	US-10-131-8268A-493
16	14	63.6	19	8	US-11-101-244-280569
C 17	14	63.6	19	8	US-11-101-244-993260
18	14	63.6	19	9	US-11-083-784-280569
C 19	14	63.6	19	9	US-11-083-784-993260
C 20	14	63.6	340000	7	US-11-102-978-3
C 21	13.8	62.7	3704	1	US-10-860-097-23
C 22	13.8	62.7	55763	1	US-10-372-766-1
C 23	13.6	61.8	1336	9	US-11-012-762-43

#### ALIGNMENTS

##### RESULT 1

US-11-101-244-280543  
; Sequence 280543, Application US/11101244  
; Publication No. US20050246794A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmacon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/101,244  
; CURRENT FILING DATE: 2005-04-07  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 280543  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-101-244-280543

Query Match 68.2%; Score 15; DB 8; Length 19;  
Best Local Similarity 73.3%; Pred. No. 1.3e+02;  
Matches 11; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTA 16

DB 5 CAUGCUCACACAUUA 19

##### RESULT 2

US-11-083-784-280543  
; Sequence 280543, Application US/11083784  
; Publication No. US20050245475A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmacon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William

Sequence 3, Appli  
Sequence 5, Appli  
Sequence 45, Appl  
Sequence 680961,  
Sequence 680961,  
Sequence 3, Appli  
Sequence 36721, A  
Sequence 56146, A  
Sequence 56648, A  
Sequence 56749, A  
Sequence 56849, A  
Sequence 144308,  
Sequence 144409,  
Sequence 799136,  
Sequence 876965,  
Sequence 1103897,  
Sequence 1132646,  
Sequence 1354276,  
Sequence 1368099,  
Sequence 1504115,  
Sequence 36721, A  
Sequence 56146, A

```
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 134990US
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR APPLICATION NUMBER: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 280543
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-280543
```

```
Query Match      68.2%; Score 15; DB 9; Length 19;
Best Local Similarity 73.3%; Pred. No. 1.3e+02;
Matches 11; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 CATGCTCACACATTA 16
         ||:|||||:|
Db      5 CAUGCUCACAUUA 19
```

```
RESULT 3
US-11-101-244-266902/c
; Sequence 266902, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 134990US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 266902
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-266902
```

```
Query Match      67.3%; Score 14.8; DB 8; Length 19;
Best Local Similarity 88.9%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      1 TCATGCTCACACATTAAC 18
         ||| ||||| |||||
Db      18 TGATGCTCACATATTAAAC 1
```

```
RESULT 4
US-11-101-244-1110410
; Sequence 1110410, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
```

```
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 134990US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 1110410
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-1110410
```

```
Query Match      67.3%; Score 14.8; DB 8; Length 19;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 12; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      5 GCTCACACATTAACGTGA 22
         ||:|||||:|
Db      1 GCUCACACAUUACUGUA 18
```

```
RESULT 5
US-11-083-784-266902/c
; Sequence 266902, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 134990US
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR APPLICATION NUMBER: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 266902
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-266902
```

```
Query Match      67.3%; Score 14.8; DB 9; Length 19;
Best Local Similarity 88.9%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      1 TCATGCTCACACATTAAC 18
         ||| ||||| |||||
Db      18 TGATGCTCACATATTAAAC 1
```

```
RESULT 6
US-11-083-784-1110410
; Sequence 1110410, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
```

```
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR APPLICATION NUMBER: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 1110410
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-1110410
```

```
Query Match 67.3%; Score 14.8; DB 9; Length 19;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 12; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 5 GCTCACACATTAACCTGGA 22
   ||:|||||: ||:|
Db 1 GCUCACACAUUUCUGUA 18
```

## RESULT 7

```
US-11-101-244-119918
; Sequence 119918, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmoon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 119918
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-119918
```

```
Query Match 65.5%; Score 14.4; DB 8; Length 19;
Best Local Similarity 68.8%; Pred. No. 2.4e+02;
Matches 11; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 7 TCACACATTAACCTGGA 22
   ||:|||||: ||:|
Db 3 UCACACAUUUCUGUA 18
```

## RESULT 8

```
US-11-101-244-993282/c
; Sequence 993282, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmoon, Inc.
; APPLICANT: Khvorova, Anastasia
```

```
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 993282
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-993282
```

```
Query Match 65.5%; Score 14.4; DB 8; Length 19;
Best Local Similarity 93.8%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 6 CTCACACATTAACCTG 21
   ||:|||||: ||:|
Db 16 CACACACATTAACCTG 1
```

## RESULT 9

```
US-11-101-244-1048490
; Sequence 1048490, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmoon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 1048490
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-1048490
```

```
Query Match 65.5%; Score 14.4; DB 8; Length 19;
Best Local Similarity 62.5%; Pred. No. 2.4e+02;
Matches 10; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 TCATGCTCACACATTA 16
   ||:|||||: ||:|
Db 4 UCAUGCUCACACAUUA 19
```

## RESULT 10

```
US-11-083-784-119918
; Sequence 119918, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmoon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
```

; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/083,784  
; CURRENT FILING DATE: 2005-03-18  
; PRIOR FILING DATE: 2003-11-14  
; PRIOR APPLICATION NUMBER: US/10/714,333  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 119918  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-083-784-119918

Query Match 65.5%; Score 14.4; DB 9; Length 19;  
Best Local Similarity 68.8%; Pred. No. 2.4e+02;  
Matches 11; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 7 TCACACATTAACCTGGA 22  
:|||||:|||||  
Db 3 UCAGACAUUACUGGA 18

RESULT 11  
US-11-083-784-993282/c  
; Sequence 993282, Application US/11083784  
; Publication No. US20050245475A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmacon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/083,784  
; CURRENT FILING DATE: 2005-03-18  
; PRIOR FILING DATE: 2003-11-14  
; PRIOR APPLICATION NUMBER: US/10/714,333  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 993282  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-083-784-993282

Query Match 65.5%; Score 14.4; DB 9; Length 19;  
Best Local Similarity 93.8%; Pred. No. 2.4e+02;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 CTCACACATTAACCTGG 21  
||||||  
Db 16 CACACACATTAACCTGG 1

RESULT 12  
US-11-083-784-1048490  
; Sequence 1048490, Application US/11083784  
; Publication No. US20050245475A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmacon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/083,784  
; CURRENT FILING DATE: 2005-03-18  
; PRIOR FILING DATE: 2003-11-14  
; PRIOR APPLICATION NUMBER: US/10/714,333  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 1048490  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-083-784-1048490

; APPLICANT: Dharmacon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/083,784  
; CURRENT FILING DATE: 2005-03-18  
; PRIOR FILING DATE: US/10/714,333  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-11-14  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 1048490  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-083-784-1048490

Query Match 65.5%; Score 14.4; DB 9; Length 19;  
Best Local Similarity 62.5%; Pred. No. 2.4e+02;  
Matches 10; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCATGCTCACACATTA 16  
:|||||:|||||  
Db 4 UCAUGCUCAACAUUA 19

RESULT 13  
US-11-101-244-876983/c  
; Sequence 876983, Application US/11101244  
; Publication No. US20050246794A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmacon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/101,244  
; CURRENT FILING DATE: 2005-04-07  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 876983  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-101-244-876983

Query Match 64.5%; Score 14.2; DB 8; Length 19;  
Best Local Similarity 84.2%; Pred. No. 2.9e+02;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTAACCTG 20  
|||||  
Db 19 CATTCCTCAACATTTACTG 1

RESULT 14  
US-11-083-784-876983/c  
; Sequence 876983, Application US/11083784  
; Publication No. US20050245475A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmacon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/101,244  
; CURRENT FILING DATE: 2005-04-07  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 876983  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-101-244-876983

GENERAL INFORMATION:  
APPLICANT: Dharmacon, Inc.  
APPLICANT: Khvorova, Anastasia  
APPLICANT: Reynolds, Angela  
APPLICANT: Leake, Devin  
APPLICANT: Marshall, William  
APPLICANT: Scaringe, Stephen  
TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
FILE REFERENCE: 13499US  
CURRENT APPLICATION NUMBER: US/11/083,784  
PRIOR FILING DATE: 2005-03-18  
PRIOR APPLICATION NUMBER: US/10/714,333  
PRIOR FILING DATE: 2003-11-14  
PRIOR APPLICATION NUMBER: 60/502,050  
PRIOR FILING DATE: 2003-09-10  
PRIOR APPLICATION NUMBER: 60/426,137  
PRIOR FILING DATE: 2002-11-14  
NUMBER OF SEQ ID NOS: 1591911  
SOFTWARE: Proprietary  
SEQ ID NO 876983  
LENGTH: 19  
TYPE: RNA  
ORGANISM: Homo sapiens  
US-11-083-784-876983

Query Match 64.5%; Score 14.2; DB 9; Length 19;  
Best Local Similarity 84.2%; Pred. No. 2.9e+02;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTAACTG 20  
DB 19 CATTTCAACATTACTG 1

RESULT 15  
US-10-131-826A-493/c  
Sequence 493, Application US/10131826A  
Publication No. US20050245730A1  
GENERAL INFORMATION:  
APPLICANT: Baker, Kevin P.  
APPLICANT: Beresini, Maureen  
APPLICANT: DeForge, Laura  
APPLICANT: Desnoyers, Luc  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerritsen, Mary E.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Sherwood, Steven  
APPLICANT: Smith, Victoria  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Watanabe, Colin K  
APPLICANT: Wood, William  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P3330R1C128  
CURRENT APPLICATION NUMBER: US/10/131,826A  
CURRENT FILING DATE: 2002-04-24  
PRIOR FILING DATE: 2002-04-24  
PRIOR APPLICATION NUMBER: 60/049911  
PRIOR FILING DATE: 1997-06-18  
PRIOR APPLICATION NUMBER: 60/056974  
PRIOR FILING DATE: 1997-08-26  
PRIOR APPLICATION NUMBER: 60/059113  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059115  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059117  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059122  
PRIOR FILING DATE: 1997-09-17

PRIOR APPLICATION NUMBER: 60/059184  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059263  
PRIOR FILING DATE: 1997-09-18  
PRIOR APPLICATION NUMBER: 60/059352  
PRIOR FILING DATE: 1997-09-19  
PRIOR APPLICATION NUMBER: 60/059588  
PRIOR FILING DATE: 1997-09-19  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 550  
SEQ ID NO 493  
LENGTH: 610  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-131-826A-493

Query Match 64.5%; Score 14.2; DB 1; Length 610;  
Best Local Similarity 84.2%; Pred. No. 2.6e+02;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 TGCTCACACATTAACTGGA 22  
DB 379 TGCTTCACATTACTGGA 361

RESULT 16  
US-11-101-244-280569  
Sequence 280569, Application US/11101244  
Publication No. US20050246794A1  
GENERAL INFORMATION:  
APPLICANT: Dharmacon, Inc.  
APPLICANT: Khvorova, Anastasia  
APPLICANT: Reynolds, Angela  
APPLICANT: Leake, Devin  
APPLICANT: Marshall, William  
APPLICANT: Scaringe, Stephen  
TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
FILE REFERENCE: 13499US  
CURRENT APPLICATION NUMBER: US/11/101,244  
CURRENT FILING DATE: 2005-04-07  
PRIOR APPLICATION NUMBER: 60/502,050  
PRIOR FILING DATE: 2003-09-10  
PRIOR APPLICATION NUMBER: 60/426,137  
PRIOR FILING DATE: 2002-11-14  
NUMBER OF SEQ ID NOS: 1591911  
SOFTWARE: Proprietary  
SEQ ID NO 280569  
LENGTH: 19  
TYPE: RNA  
ORGANISM: Homo sapiens  
US-11-101-244-280569

Query Match 63.6%; Score 14; DB 8; Length 19;  
Best Local Similarity 71.4%; Pred. No. 3.5e+02;  
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 CATGCTCACACATT 15  
DB 6 CAUGCUCACACAUU 19

RESULT 17  
US-11-101-244-993260/c  
Sequence 993260, Application US/11101244  
Publication No. US20050246794A1  
GENERAL INFORMATION:  
APPLICANT: Dharmacon, Inc.  
APPLICANT: Khvorova, Anastasia  
APPLICANT: Reynolds, Angela  
APPLICANT: Leake, Devin  
APPLICANT: Marshall, William  
APPLICANT: Scaringe, Stephen  
TITLE OF INVENTION: Functional and Hyperfunctional siRNA

```
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 993260
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-993260

Query Match      63.6%; Score 14; DB 8; Length 19;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      8 CACACATTAACCTGG 21
      |||||
Db      18 CACACATTAACCTGG 5

RESULT 18
US-11-083-784-280569
; Sequence 280569, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR APPLICATION NUMBER: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 280569
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-280569

Query Match      63.6%; Score 14; DB 9; Length 19;
Best Local Similarity 71.4%; Pred. No. 3.5e+02;
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY      2 CATGCTCACACATT 15
      ||:|||||:
Db      6 CAUGCUCACACAUU 19

RESULT 19
US-11-083-784-993260/c
; Sequence 993260, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
```

```
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR APPLICATION NUMBER: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 993260
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-993260

Query Match      63.6%; Score 14; DB 9; Length 19;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      8 CACACATTAACCTGG 21
      |||||
Db      18 CACACATTAACCTGG 5

RESULT 20
US-11-102-978-3/c
; Sequence 3, Application US/11102978
; Publication No. US20050250142A1
; GENERAL INFORMATION:
; APPLICANT: University of Utah Technology Transfer Office
; APPLICANT: University of Utah Research Foundation
; TITLE OF INVENTION: Diagnosis and Treatment of Herpes Simplex Virus Disease
; FILE REFERENCE: 0274-5537.1US
; CURRENT APPLICATION NUMBER: US/11/102,978
; CURRENT FILING DATE: 2005-04-11
; PRIOR APPLICATION NUMBER: PCT/US2003/033152
; PRIOR FILING DATE: 2003-10-18
; PRIOR APPLICATION NUMBER: 60/419,576
; PRIOR FILING DATE: 2002-10-18
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 340000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: exon
; LOCATION: (56948)..(57115)
; OTHER INFORMATION: C21orf34 exon
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (80006)..(81089)
; OTHER INFORMATION: Gene VDACC2P; voltage-dependent anion channel isoform 2 pseudogen
; FEATURE:
; NAME/KEY: exon
; LOCATION: (167308)..(167438)
; OTHER INFORMATION: C21orf34 exon
; FEATURE:
; NAME/KEY: exon
; LOCATION: (216732)..(216833)
; OTHER INFORMATION: C21orf34 exon
US-11-102-978-3

Query Match      63.6%; Score 14; DB 7; Length 340000;
Best Local Similarity 77.3%; Pred. No. 2e+02;
Matches 17; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 TCATGCTCACACATTAACCTGGA 22
      |||||
Db      26982 TCATGCTCAATCATTTAATAGA 26961
```



```
; APPLICANT: Andersen, Carsten
; APPLICANT: Fuglsang, Claus Crone
; TITLE OF INVENTION: Alpha-amylase mutants with altered properties
; FILE REFERENCE: 10062.200-US
; CURRENT APPLICATION NUMBER: US/10/630,203
; CURRENT FILING DATE: 2003-07-29
; PRIOR APPLICATION NUMBER: US/09/918,543
; PRIOR FILING DATE: 2001-07-31
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 3
; LENGTH: 1455
; TYPE: DNA
; ORGANISM: Bacillus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1455)
; OTHER INFORMATION:
US-10-630-203-3

Query Match      61.8%; Score 13.6; DB 1; Length 1455;
Best Local Similarity 80.0%; Pred. No. 4.7e+02;
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      2 CATGCTCACACATTAACTGG 21
      ||| ||| ||| ||| ||| ||| |||
Db      452 CATACTCAGACTTTAAATGG 471

RESULT 25
US-11-103-037-5
; Sequence 5, Application US/11103037
; Publication No. US20050250664A1
; GENERAL INFORMATION:
; APPLICANT: Biscard-Frantzen, Henrik
;              Borchert, Torben Vedel
;              Svendsen, Allan
; TITLE OF INVENTION: AMYLASE VARIANTS
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Novo Nordisk of North America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/11/103,037
; FILING DATE: 11-Apr-2005
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/902,188
; FILING DATE: 10-Jul-2001
; APPLICATION NUMBER: 09/354,191
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4318.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1455 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
```

```
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-11-103-037-5

Query Match      61.8%; Score 13.6; DB 7; Length 1455;
Best Local Similarity 80.0%; Pred. No. 4.7e+02;
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      2 CATGCTCACACATTAACTGG 21
      ||| ||| ||| ||| ||| ||| |||
Db      452 CATACTCAGACTTTAAATGG 471

RESULT 26
US-11-012-762-45/c
; Sequence 45, Application US/11012762
; Publication No. US20050244815A1
; GENERAL INFORMATION:
; APPLICANT: Georgia State University Research Foundation, Inc.
; TITLE OF INVENTION: Compositions and Methods for Viral Resistance Genes
; FILE REFERENCE: GSUI.PCT
; CURRENT APPLICATION NUMBER: US/11/012,762
; CURRENT FILING DATE: 2004-12-15
; PRIOR APPLICATION NUMBER: PCT/US03/19300
; PRIOR FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US 60/390,046
; PRIOR FILING DATE: 2002-06-19
; NUMBER OF SEQ ID NOS: 133
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 45
; LENGTH: 1559
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (6)..(1076)
US-11-012-762-45

Query Match      61.8%; Score 13.6; DB 9; Length 1559;
Best Local Similarity 80.0%; Pred. No. 4.7e+02;
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      2 CATGCTCACACATTAACTGG 21
      ||| ||| ||| ||| ||| ||| |||
Db      297 CTGTGCGCCACTTTAACTGG 278

RESULT 27
US-11-101-244-680961/c
; Sequence 680961, Application US/1101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmoon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 680961
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-680961
```



Query Match 60.9%; Score 13.4; DB 8; Length 19;  
Best Local Similarity 93.3%; Pred. No. 6.3e+02;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCATGCTCACACATT 15  
|||||  
DB 19 TCATGCTCACACCTT 5

## RESULT 28

US-11-083-784-680961/c  
; Sequence 680961, Application US/11083784  
; Publication No. US20050245475A1

; GENERAL INFORMATION:

; APPLICANT: Dharmacon, Inc.

; APPLICANT: Khvorova, Anastasia

; APPLICANT: Reynolds, Angela

; APPLICANT: Leake, Devin

; APPLICANT: Marshall, William

; APPLICANT: Scaringe, Stephen

; TITLE OF INVENTION: Functional and Hyperfunctional siRNA

; FILE REFERENCE: 13499US

; CURRENT APPLICATION NUMBER: US/11/083,784

; CURRENT FILING DATE: 2005-03-18

; PRIOR APPLICATION NUMBER: US/10/714,333

; PRIOR FILING DATE: 2003-11-14

; PRIOR APPLICATION NUMBER: 60/502,050

; PRIOR FILING DATE: 2003-09-10

; PRIOR APPLICATION NUMBER: 60/426,137

; PRIOR FILING DATE: 2002-11-14

; NUMBER OF SEQ ID NOS: 1591911

; SOFTWARE: Proprietary

; SEQ ID NO 680961

; LENGTH: 19

; TYPE: RNA

; ORGANISM: Homo sapiens

US-11-083-784-680961

Query Match 60.9%; Score 13.4; DB 9; Length 19;  
Best Local Similarity 93.3%; Pred. No. 6.3e+02;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCATGCTCACACATT 15  
|||||  
DB 19 TCATGCTCACACCTT 5

## RESULT 29

US-10-663-794-3/c

; Sequence 3, Application US/10663794

; Publication No. US20050244833A1

; GENERAL INFORMATION:

; APPLICANT: WEI, Ming-Hui et al.

; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC

; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES

; TITLE OF INVENTION: THEREOF

; FILE REFERENCE: CL001164CIP-DIV II

; CURRENT APPLICATION NUMBER: US/10/663,794

; CURRENT FILING DATE: 2003-09-17

; NUMBER OF SEQ ID NOS: 6

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 3

; LENGTH: 48763

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-663-794-3

Query Match 60.9%; Score 13.4; DB 1; Length 48763;  
Best Local Similarity 93.3%; Pred. No. 4.7e+02;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 CTCACACATTAACTG 20

Db 11635 CTCACACATTAACTG 11621  
|||||

## RESULT 30

US-11-101-244-36721

; Sequence 36721, Application US/11101244

; Publication No. US20050246794A1

; GENERAL INFORMATION:

; APPLICANT: Dharmacon, Inc.

; APPLICANT: Khvorova, Anastasia

; APPLICANT: Reynolds, Angela

; APPLICANT: Leake, Devin

; APPLICANT: Marshall, William

; APPLICANT: Scaringe, Stephen

; TITLE OF INVENTION: Functional and Hyperfunctional siRNA

; FILE REFERENCE: 13499US

; CURRENT APPLICATION NUMBER: US/11/101,244

; CURRENT FILING DATE: 2005-04-07

; PRIOR APPLICATION NUMBER: 60/502,050

; PRIOR FILING DATE: 2003-09-10

; PRIOR APPLICATION NUMBER: 60/426,137

; PRIOR FILING DATE: 2002-11-14

; NUMBER OF SEQ ID NOS: 1591911

; SOFTWARE: Proprietary

; SEQ ID NO 36721

; LENGTH: 19

; TYPE: RNA

; ORGANISM: Homo sapiens

US-11-101-244-36721

Query Match 60.0%; Score 13.2; DB 8; Length 19;  
Best Local Similarity 61.1%; Pred. No. 7.7e+02;  
Matches 11; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 5 GCTCACACATTAACTGGA 22  
|||:||||:|:|  
DB 1 GCUCUCACCUAUCUGGA 18

## RESULT 31

US-11-101-244-56146

; Sequence 56146, Application US/11101244

; Publication No. US20050246794A1

; GENERAL INFORMATION:

; APPLICANT: Dharmacon, Inc.

; APPLICANT: Khvorova, Anastasia

; APPLICANT: Reynolds, Angela

; APPLICANT: Leake, Devin

; APPLICANT: Marshall, William

; APPLICANT: Scaringe, Stephen

; TITLE OF INVENTION: Functional and Hyperfunctional siRNA

; FILE REFERENCE: 13499US

; CURRENT APPLICATION NUMBER: US/11/101,244

; CURRENT FILING DATE: 2005-04-07

; PRIOR APPLICATION NUMBER: 60/502,050

; PRIOR FILING DATE: 2003-09-10

; PRIOR APPLICATION NUMBER: 60/426,137

; PRIOR FILING DATE: 2002-11-14

; NUMBER OF SEQ ID NOS: 1591911

; SOFTWARE: Proprietary

; SEQ ID NO 56146

; LENGTH: 19

; TYPE: RNA

; ORGANISM: Homo sapiens

US-11-101-244-56146

Query Match 60.0%; Score 13.2; DB 8; Length 19;  
Best Local Similarity 72.2%; Pred. No. 7.7e+02;  
Matches 13; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 GCTCACACATTAACTGGA 22  
||| |||||:|:|

Db 1 GCCCACAAGCAACUGGA 18

## RESULT 32

US-11-101-244-56648  
; Sequence 56648, Application US/11101244  
; Publication No. US20050246794A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmacon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/101,244  
; CURRENT FILING DATE: 2005-04-07  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 56648  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-101-244-56648

Query Match 60.0%; Score 13.2; DB 8; Length 19;  
Best Local Similarity 72.2%; Pred. No. 7.7e+02;  
Matches 13; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 GCTCACACATTAACUGGA 22

Db 1 GCCCACAAGCAACUGGA 18

## RESULT 33

US-11-101-244-56749  
; Sequence 56749, Application US/11101244  
; Publication No. US20050246794A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmacon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/101,244  
; CURRENT FILING DATE: 2005-04-07  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 56749  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-101-244-56749

Query Match 60.0%; Score 13.2; DB 8; Length 19;  
Best Local Similarity 72.2%; Pred. No. 7.7e+02;  
Matches 13; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 GCTCACACATTAACUGGA 22

Db 1 GCCCACAAGCAACUGGA 18

## RESULT 34

US-11-101-244-56849  
; Sequence 56849, Application US/11101244  
; Publication No. US20050246794A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmacon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/101,244  
; CURRENT FILING DATE: 2005-04-07  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 56849  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-101-244-56849

Query Match 60.0%; Score 13.2; DB 8; Length 19;  
Best Local Similarity 72.2%; Pred. No. 7.7e+02;  
Matches 13; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 GCTCACACATTAACUGGA 22

Db 1 GCCCACAAGCAACUGGA 18

## RESULT 35

US-11-101-244-144308/c  
; Sequence 144308, Application US/11101244  
; Publication No. US20050246794A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmacon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/101,244  
; CURRENT FILING DATE: 2005-04-07  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 144308  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-101-244-144308

Query Match 60.0%; Score 13.2; DB 8; Length 19;  
Best Local Similarity 83.3%; Pred. No. 7.7e+02;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CATGCTCACACATTAAC 19

Db 19 CATTACACAGATTAAC 2

RESULT 36

```

US-11-101-244-144409/c
; Sequence 144409, Application US/111011244
; Publication NO. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hype
; FILE REFERENCE: 134990S
; CURRENT APPLICATION NUMBER: US/11/101,2
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 144409
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-144409

```

Query Match 60.0%; Score 13.2; DB 8; Length 19;  
Best Local Similarity 83.3%; Pred. No. 7.7e+02;  
Matches 15; Conservative 0; Mismatches 3; Indels

Qy 2 CATGCTCACACATTAAC 19  
||| | |||| |||||  
Db 19 CATTACACAGATTAAC 2

RESULT 37

```

RESULT 37
US-11-101-244-799136
; Sequence 799136, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyper
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,2
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 799136
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-799136

```

Query Match 60.0%; Score 13.2; DB 8; Length 19;  
Best Local Similarity 72.2%; Pred. No. 7.7e+02;  
Matches 13; Conservative 2; Mismatches 3; Indels

Qy 5 GCTCACACATTAACTGGA 22  
||:|||||:|||||  
Db 1 GTCACAGAUACAACAGGA 18

RESULT 38

```

US-11-101-244-876965/c
; Sequence 876965, Application US/11101244
; Publication No. US20050246794A1
GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scarsinge, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13495US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502, 050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426, 137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 876965
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-876965

```

Query Match	60.0%	Score 13.2;	DB 8;	Length 19;
Best Local Similarity	83.3%	Pred. No. 7.7e+02;		
Matches 15; Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0;

Qy            4 TGCTCACACATTTAACTGG 21  
               | ||||| ||||| |||||  
Db            19 TTCTCAAACATTTACTGG 2

RESULT 39

```

RESUL1 39
US-11-101-244-1103897
; Sequence 1103897, Application US/11101244
; Publication NO. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 134990US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 1103897
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-1103897

```

Query Match 60.0%; Score 13.2; DB 8; Length 19;  
Best Local Similarity 61.1%; Pred. No. 7.7e+02;  
Matches 11; Conservative 4; Mismatches 3; Indels

Qy 1 TCATGCTCACACATTAAAC 18  
: || || : || || : || ||  
Db 2 UCAAGCUCUUAACAUTAAAC 19

## RESULT 40

```

US-11-101-244-1132646/c
; Sequence 1132646, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 1132646
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-1132646

```

```

Query Match      60.0%; Score 13.2; DB 8; Length 19;
Best Local Similarity 83.3%; Pred. No. 7.7e+02;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      4 TGCTCACACATTAACTGG 21
Db      18 TACTCAGATATTAAC TGG 1

```

Search completed: November 21, 2005, 05:22:42  
Job time : 307.457 secs